

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:10 ; Search time 71 Seconds
(without alignments)

2536.368 Million cell updates/sec

Title: US-10-009-823a-1

Perfect score: 2586

Sequence: 1 MMGSLFVGATGMKTHSTGLG.....NSKSVTTADTLMQALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A: Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2484	96.1	486	4	AAB36552
2	538	20.8	120	2	AAW16580
3	529.5	20.5	414	6	ABU20217
4	528.5	20.4	471	7	ABO70780
5	527.5	20.4	462	6	ABJ18752
6	506.5	19.6	413	6	ABU22839
7	491.5	19.0	402	8	ADK13784
8	484.5	18.7	411	5	ABB49883
9	483	18.7	718	2	AAW80991
10	475.5	18.4	412	6	ADF07897
11	448.5	17.3	405	6	ABM68268
12	376.5	14.6	565	4	ABG24709
13	307	11.9	269	7	ABO70969
14	296	11.4	261	6	ABM68266
15	288	11.1	271	2	AAW20373
16	288	11.1	271	2	AAW24645
17	288	11.1	279	2	AAW20863
18	286	11.1	265	7	ADF07825
19	237.5	9.2	263	3	AAV51377
20	223.5	8.6	263	3	AAW73073
21	176	6.8	2736	7	ASO81481
22	173.5	6.7	1095	3	AAW01847
23	173.5	6.7	1536	2	AAW01844
24	173.5	6.7	1536	2	AAW01844
25	173.5	6.7	1536	2	AAW63505

26	173.5	6.7	1536	3	AAW01846
27	172.5	6.7	3705	6	ABU50020
28	167.5	6.5	1536	2	AAW30293
29	166	6.4	2468	6	ABU38411
30	166	6.4	2468	6	ABP59933
31	165.5	6.4	124	2	AAW20361
32	165	6.4	5291	7	ADC01014
33	162	6.3	1954	8	ADF83256
34	162	6.3	1954	8	ADK13634
35	160	6.2	1230	2	AAW98275
36	160	6.2	1230	2	AAW98275
37	156.5	6.1	3073	6	ABU30969
38	155	6.0	270	3	AAW73072
39	155	5.9	270	3	AAW73072
40	151.5	5.9	2204	6	ABU36440
41	149.5	5.8	1063	5	ABW54168
42	149.5	5.8	1180	3	AAW01845
43	149.5	5.8	1188	3	AAW01844
44	148.5	5.7	273	2	AAW20658
45	147	5.7	1468	7	ADD48744

ALIGNMENTS

RESULT 1

AAB36552
ID AAB36552 standard; protein; 486 AA.

XX AAB36552;

XX 07-MAR-2001 (first entry)

XX Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.

XX Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;
XX intestinal disease; immunogenic; diagnosis; antibacterial; swine; pig;
XX infection; detection; identification.

XX Lawsonia intracellularis.

XX W0200069904-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-AU000437.

XX 13-MAY-1999; 99US-0133973P.

XX (PFIZ) PFIZER PROD INC.
XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
XX (PIGR-) PIG RES & DEV CORP.

XX Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;
XX Ankenbauer RG;

XX WPI; 2001-016210/02.

XX N-PSDB; AAC88037.

XX New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody, in
XX useful in vaccines and diagnosis of Lawsonia infections, particularly in
XX swine.

XX Claim 13; Page 87-90; 97pp; English.

XX The present sequence is the Lawsonia intracellularis flagellar hook
XX protein FlgE. The present invention describes an isolated or recombinant
XX polypeptide (I) that comprises, mimics or cross-reacts with a B- or T-
XX cell epitope of a FlgE (flagellar hook) polypeptide from a Lawsonia spp.
XX (I) has antibacterial activity, and induces a specific humoral immune
XX response. (I) are used as antigens in vaccines to prevent or treat
XX infection by Lawsonia, in birds and animals, especially pigs, to raise
XX specific antibodies (Ab) and to detect past or present infection. Ab are

CC also useful in diagnosis, to detect *L. intracellularis* or immunologically
 CC cross-reactive species, also for identification of epitopes in FlgE.
 CC Vectors that contain nucleic acids (ii) encoding (i) are also useful in
 CC genetic vaccines, and fragments of (ii) are useful as primers or probes
 CC for detecting *L. intracellularis* or related microorganisms, in
 CC hybridisation or amplification assays
 XX
 SQ Sequence 486 AA;

Query Match 36.1%; Score 2484; DB 4; Length 486;
 Best Local Similarity 96.8%; Pred. No. 9.7e-178;
 Matches 486; Conservative 0; Mismatches 0; Indels 16; Gaps 1;
 QY 1 MMGSLFIGATGKTHSTGLTGVNNANANTIGYKQQVWFQDLFSQDLAIGTSGSQGN 60
 DB 1 MMGSLFIGATGKTHSTGLTGVNNANANTIGYKQQVWFQDLFSQDLAIGTSGSQGN 60
 QY 61 QAGMGAGVSVRTFTQGAPEFGNSVTDLAIGGKGFQVTLKDYHYTRAGNFRFTQDGF 120
 DB 61 QAGMGAGVSVRTFTQGA-----FFQVTLKDYHYTRAGNFRFTQDGF 104
 QY 121 LNDPSGFTLMGSRISNNPNKXETLEPIQLDNDFTVAKSPAKTALNAVNLGSDTDK 180
 DB 105 LNDPSGFTLMGSRISNNPNKXETLEPIQLDNDFTVAKSPAKTALNAVNLGSDTDK 164
 QY 181 TQSEANYPALLESKNGKNGTPTPSTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGSK 240
 DB 165 TQSEANYPALLESKNGKNGTPTPSTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGSK 224
 QY 241 TPEYLVAMPSPEDGSAAGTDSAGLLMGSTMTFSSNGELKNTAFTPGSATKDLNAWQP 300
 DB 225 TPEYLVAMPSPEDGSAAGTDSAGLLMGSTMTFSSNGELKNTAFTPGSATKDLNAWQP 284
 QY 301 APLVNLGPOFSANFVGAGIQLTLDFGIKSOQNMWAGAPASAAAGTIDIGKLPNMPIQT 360
 DB 285 APLVNLGPOFSANFVGAGIQLTLDFGIKSOQNMWAGAPASAAAGTIDIGKLPNMPIQT 344
 QY 361 SSGNSTARGSSSTRYSQDGPQDLDVDTITSEGLQKGYKSNQVDFNIPLARFTS 420
 DB 345 SSGNSTARGSSSTRYSQDGPQDLDVDTITSEGLQKGYKSNQVDFNIPLARFTS 404
 QY 421 EDGLREGNNHYSATLDSGGPEFGLPSTSNYKLSVNOLETNSVDMGEMVNMIIIOGPF 480
 DB 405 EDGLREGNNHYSATLDSGGPEFGLPSTSNYKLSVNOLETNSVDMGEMVNMIIIOGPF 464
 QY 481 QMNSKSVTTADTMLOKALELKR 502
 DB 465 QMNSKSVTTADTMLOKALELKR 486

RESULT 2
 AAU16680
 ID AAU16680 standard; protein; 120 AA.
 XX
 AC AAU16680;
 XX
 DT 20-AUG-1997. (first entry)
 XX
 DE Lawsonia intracellularis flagellar basal body rod-like protein.
 XX
 KW Intestinal disease; porcine proliferative enteropathy; vaccine; GroEL;
 immunotherapy; antibody; diagnosis; flagellar basal body rod protein.
 XX
 OS Lawsonia intracellularis.
 XX
 EH Key Location/Qualifiers
 FT Misc-difference 5
 FT /note= "residue 5 corresponds to degenerate codon (GNT)
 in the nucleotide sequence"
 FT
 FT Misc-difference 12
 FT /note= "residue 12 corresponds to an in-frame stop codon
 in the nucleotide sequence"
 FT
 FT Misc-difference 13

FT /note= "residue 13 corresponds to degenerate codon (NAC)
 in the nucleotide sequence"
 FT Misc-difference 16
 FT /note= "residue 16 corresponds to degenerate codon (GNT)
 in the nucleotide sequence"
 XX
 PN W09720050-A1.
 XX
 PD 05-JUN-1997.
 XX
 XX 29-NOV-1996; 96WO-AU000767.
 XX
 PR 30-NOV-1995; 95AU-00006910.
 XX
 PR 30-NOV-1995; 95AU-00006911.
 XX
 XX (DARA-) DARATECH PTY LTD.
 PA (PIGR-) FIG RES & DEV CORP.
 XX
 XX Panaccio M, Hasse D;
 XX
 XX WPI; 1997-310605/28.
 DR N-PSDB; AAT69204.
 XX
 XX Vaccine for treating or preventing Lawsonia intracellularis infection -
 especially in pigs, containing non-pathogenic form of bacterium or its
 components.
 FT
 FT
 XX
 PS Claim 26; Page 51-52; 94pp; English.
 XX
 CC A polypeptide (AAU16680) of Lawsonia intracellularis shows sequence
 similarity to flagellar basal body rod protein. Its amino acid sequence
 was deduced from a DNA molecule (AAT69204) obt'd. by immunoscreening a *L.*
 intracellularis library using sera from vaccinated pigs. Polypeptides
 (AAU16678-85) of *L. intracellularis* are vaccine candidates for protection
 of birds and animals against intestinal diseases, esp. protection of pigs
 against porcine proliferative enteropathy. They can be administered as
 recombinant polypeptides or expressed as recombinant vaccines utilising
 bacterial, fungal or viral vectors. Antibodies raised against the
 polypeptides may be useful in immunotherapy, diagnosis and detection
 SQ Sequence 120 AA;
 Query Match 20.8%; Score 538; DB 2; Length 120;
 Best Local Similarity 89.2%; Pred. No. 1.7e-32;
 Matches 107; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
 QY 4 SLFIGATGKTHSTGLTGVNNANANTIGYKQQVWFQDLFSQDLAIGTSGSQGNQAG 63
 DB 1 SLFIGATGKTHSTGLTGVNNANANTIGYKQQVWFQDLFSQDLAIGTSGSQGNQAG 60
 QY 64 MGAQVGSVRTFTQGAPEFGNSVTDLAIGGKGFQVTLKDYHYTRAGNFRFTQDGFND 123
 DB 61 MGAQVGSVRTFTQGAPEFGNSVTDPAIGGKGFQVTLKDYHYTRAGNFRFTQDGFND 120
 RESULT 3
 ABU20217
 ID ABU20217 standard; protein; 414 AA.
 XX
 AC ABU20217;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #5744.
 XX
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Borrelia cepacia.
 XX
 PN W0200277183-A2.
 XX
 PD 03-OCT-2002.

CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html

```

Query Match      20.4%; Score 528.5; DB 7; Length 471;
Best Local Similarity 28.5%; Pred. No. 5.9e-31;
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;

QY      4 SLFICATGKTHSTGLGTVSNNTANANTICYKKOQVVFDLPSODLAIGSTGSQCQNAG 63
Db      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
11 SNIIGLSGIQAASSGLNVTGNNTANAGTVGFKOSRAEFADVYAASV-LGS-----GSPNQG 65

QY     64 MGAQGVSVRTIFTQCAPEPGNSVTDLAIGCKGFQVTTLEDKVHYTRAGNFRAFTODGFLND 123
Db      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
66 SGVLLSDYSQMFKGQDNIDSTNVLDLAINGGEFVTSNNCAISYTRAGVFNTDKQDIFVD 125

QY    124 PSGFTLMGSRISNNPNTKETLEPIQLD-FNDPTVAKSPAKTSTALNAVNLGSDTOKTQ 182
Db      ::||::||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
126 NNGYRLQGYAVGPNGQLQGNGVWTDLKVERANAPOQATSSIIQQSYNLNSTLK-----P 177

QY    193 SEANPYFALLESWKNGCTPISNTSNTSYAQPMRVYDOOGNSHDITVP--DGAPST-- 237
Db      ::||::||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
178 PTVTFP-----DPSDAATYNSSSLGIYDSQGNSTWSQFFIKNEPFENATPP 225

QY   238 ---GSKTPEYLIV-AMNPSEPGSAASGTDAGLIMSGTWTFSSNGELKNMTAFDTGS--- 290
Db      ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
226 IPENSWTMKVLIDGVNPLDSPNKTP-----MSFNVTFDASGGOMTSVRA--PDGSTG 275

QY   291 -----ATKDNLNAQAPLVNGLPQFSANFVGAGIQPLTLDFGIKSQONMWAGAPASAAA 344
Db      ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
276 PGFSIDATTNVIQSPA---TGNPPTP----GTGMIPAAAD-GKTPPTYAMKATGAASG 327

QY   345 ICTDIGKLPSMMPLOTSSGNSTARNSSSTRAYS-----QDCYPCQGLDVDTITSE 395
Db      ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
328 ISFMRK-----TTQYSTAFAGSNPIQDGYTTGQLAGLEIDDT 365

QY   396 GKLOCKYSNOVDFYNIPLARFTSBEDGLRREGNHHYSATLDSGGEPGLPCTSNYGKLS 455
Db      ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
366 GVIFARYNGSKVCQOVVLNANFANTQGLTPIGKTSWVSASESEPACPRSTGLCALQ 425

QY   456 VNQLETSNVDMSRMVMNMIIRGFQPMNSKSVTTADTMLOKALELK 501
Db      ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
426 SGAAEASNVDIISNELVNLIYHONYOANAKTTOTEDAVTOTINLR 471

```

RESULT 5	
ABJ18752	
ID	ABJ18752 standard; protein; 462 AA.
XX	
XX	
AC	ABJ18752;
XX	
XX	
DT	27-FEB-2003 (first entry)
XX	
XX	
DE	Pseudomonas aeruginosa biofilm formation-related protein #16.
XX	
XX	Biofilm formation modulation; biofilm-associated disease;
KW	cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
KW	catheter-associated infection; medical device-associated infection.
XX	
XX	
CS	Pseudomonas aeruginosa.
XX	
XX	
FN	WO200285295-A2.
XX	
XX	
PD	31-OCT-2002.

PA	(HARD) HARVARD COLLEGE.
XX	
PI	Whiteley M, Bangera MG, Lory S, Greenberg EP;
XX	
DR	WFI; 2003-075601/07.
DR	N-PSDB; ABT14574.
XX	
PT	Identifying compound capable of modulating biofilm formation by
PT	bacteria/bacterial antibiotic resistance, useful for treatment of biofilm
PT	associated disease.
XX	
XX	Claim 1; Page 101-105; 154pp; English.
XX	
CC	The invention comprises a method for identifying a compound capable of
CC	modulating biofilm formation by bacteria. The method of the invention is
CC	useful for identifying a compound capable of modulating biofilm formation
CC	by bacteria or modulating bacterial antibiotic resistance. The method of
CC	the invention is also useful for diagnosing and treating a subject
CC	(especially an immunocompromised human) that is afflicted with a biofilm-
CC	associated disease or disorder, such as: cystic fibrosis; AIDS; middle
CC	ear infections; acne; periodontal disease; catheter-associated infections
CC	; and medical device-associated infections. The present amino acid
CC	sequence represents a protein that is used in the invention
XX	
XX	Sequence 462 AA;

Query Match	20.4%;	Score 527.5;	DB 6;	Length 462;
Best Local Similarity	28.5%;	Pred. No. 6.9e-31;		
Matches	150;	Conservative 84;	Mismatches 199;	Gaps 16;
QY	4	SLPFGATGMKTHSGLCTVSNNTANITGYKQOVVFODLFSQDLAIGSTGSGPNQAG	63	
DB	2	SNIGLSIGIAASSGLNVTGNNTANAGTVGFKQSRAEFADYAAVS-LGS-----GSPNQG	56	
QY	64	MGAOVGSVRTTFTQCAFEPSNVTDLAIGGKGFFOVTLDEKHYHTRAGNFRFTDQGFLLND	123	
DB	57	SGVLLSDVSQMFQKGNIDSTNSVLDLAINGNGFFVTSNNGAISYTRAGYENTDKQDFV	116	
QY	124	PSGFLLMGSRISNNPNITKKTLEPIQLD-FNDPTVAKSPAKTSTALNAVVLGSDTKTQ	182	
DB	117	NNGYRLQYAVGPNGLQNGVWDLKVERANAQPAQTSSIQOSYNLNLTK-----P	168	
QY	193	SEANPYFALLESWKNGCTPPISTNSYSAQPKRVYDQCNSHDITVYF-----DGA	237	
DB	169	PTVTVPF-----DPSDAATYNSSSSLGIYDSQGNSHTMQOFFIKNEPDPNATPP	216	
QY	238	--GSKTPEYLV-AMNPSEDSGAASGTDAGLLMSGTMTFSSNGELKKNKTAFTPGS---	290	
DB	217	IPENSWTKVLLIDGVNPLDSKNTP-----MSFNVTFDASGQMTSVRA--PDG	266	
QY	291	-----ATKDLNAWQAPLNYGLPQFSANFVAGIQPLTLDFGIKSQONMAGAPASAAA	344	
DB	267	PGFSIDATTNVIQFSPA---TGNPPTP---GTGMIPAAASD-GKTPPTPYAWNGATGAAS	318	
QY	345	IGTDIGKLPSMPTQTSNGNSTARNSSSTRYS-----QDGYPCGLVDVTTTSE	395	
DB	319	ISFPMRX-----TTOYSTAFAGSNPIDQYGTGTGQLAGLEIDDT	356	
QY	396	GKLGQKYSNSQVDFVNIPIARFTSBDGLRRRENNHYSATLDSGGPFGLPGTSGNYGKLS	455	
DB	357	GVI FARVTNGSKVQGOVVLNAPANTCGLTPTGKTSWVSSESSEGFVAGAPRSGTGLGALQ	416	
QY	456	VNQLSTSNVDMSRMNMIIRGFQFMNSKSVTTADTMLQKALEUK	501	
DB	417	SGALEASNDVISELNVLIWHQNYQANAKTIQTSDAVTQTTINLR	462	

RESULT 6
ABU22839
ID ABU22839 standard; protein; 413 AA.
XX
XX ABU22839?
XX

19-JUN-2003 (first entry)
 Protein encoded by Prokaryotic essential gene #8366.
 Antisense; prokaryotic essential gene; cell proliferation; drug design.
 Burkholderia mallei.
 W020027183-A2.
 03-OCT-2002.
 21-MAR-2002; 2002WO-US009107.
 21-MAR-2001; 2001US-00815242.
 06-SEP-2001; 2001US-00948993.
 25-OCT-2001; 2001US-0342923P.
 08-FEB-2002; 2002US-00072851.
 06-MAR-2002; 2002US-0362699P.
 (ELIT-) ELITRA PHARM INC.
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 WPI; 2003-029926/02.
 N-PSDB; ACA26709.
 New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 25; SEQ ID NO 50763; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
 the 5213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
X. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

Sequence 413 AA;
 Query Match 19.6%; Score 506.5; DB 6; Length 413;
 Best Local Similarity 27.2%; Pred. No. 2,2e-29;
 Matches 137; Conservative 81; Mismatches 180; Indels 105; Gaps 11;
 8 GATGKTHSTGLTGNNTANNTIGYKQQVVFQDLAIGTSQSGNQMGQAQ 67

Db	6	GLSLGAGASSDLDVIGNNANANTVFGKSTAQFSDMYANSVA-----SAVNNFIGITM	60
Qy	68	VGSVETITQGAPEFGNSVTDLAIGGKGFQVLTLEDKVVHYTRAGNFRFTQDGLNDPSGF	127
Db	61	LASVQQQSFSGTITSTSSLNVAINGNGFPQNSNGVTVTYSRDTGTFQDKNGYIVNSQGL	120
Qy	128	TLMSGRISNNPNKKEITLEPIQLDFNDPTVAKSPAKTSTALNAVNVNLDGSDTKTQSEANP	187
Db	121	NLMGYAANANGVINTAATVPLQA-----PTTNIAPTAT-TKITQFNL-NSQDAVPA-TTP	173
Qy	189	YFALLESWKNGCTPPISTNSYSAQPMVYDOQGNSHDITVYEDGAPSSSTGKTFEYLV	247
Db	174	F-----NYTDPTS-----YNYTTSVQVFDLGGSQNNVLFVKSATSGQWEAYAGPAG	221
Qy	248	MNPSDEGSAASCTDSAGLMSGMTTFSSNGELKNNATFTPTGSATKDLNAMQAPLNVGL	307
Db	222	KTFPTDLS-----VKFSTAGTITGTSTP-----	244
Qy	308	POFSANFVGAGIQPLTLDPGIKSQNNMAGAPASAAAGTIDIGKLPSSMPTQSSGN---	364
Db	245	-----AGVPT-----TNVQSFSTPTTIGANPON	270
Qy	365	-----STARNGSSSTRYSQDGYPGQDLVDVITTSBGLQKQKYSNSQVDFYNNIPLAR	417
Db	271	LTLDLTGITQYGGKNGINLAQDGFASGVLTTFSGADGKVTGNYNSGQSTLGIQVLAN	330
Qy	418	FTSEDLRREGNNHYSATLDSGGPFGPPTGSNYKLSVNOLETNSVDMSEKVMIIQ	477
Db	331	FNNPGLNVNNGVYVESAASGVQITSAPGSTNHGTQSGALENSNVDLTSLQVLKLTAAQ	390
Qy	478	RGFQNSKSVTTADTMLQALEL	500
Db	391	RNYQANAQTIKTQQTVDQTLINL	413

RESULT 7
 ADK13784
 ID ADK13784 standard; protein; 402 AA.
 AC ADK13784;
 XX
 DT 20-MAY-2004 (first entry)
 DE E. coli iron transport and metabolism protein SEQ ID NO:79.
 XX
 XX Escherichia coli; iron transport and metabolism protein; iron transport;
 XX iron metabolism; antibacterial.
 OS Escherichia coli.
 XX
 XX W02004018638-A2.
 PN
 PD 04-MAR-2004.
 XX
 PF 21-AUG-2003; 2003WO-US026488.
 XX
 PR 21-AUG-2002; 2002US-0405331P.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (KAPU) KAPUR V.
 PA (GADG) GADGIL M.
 XX
 PI Kapur V, Gadgil M;
 XX
 DR WPI; 2004-238974/22.
 DR N-PSDB; ADK13720.
 XX
 PT New isolated and purified iron transport and metabolism polypeptides and
 PT encoding polynucleotides, useful in identifying potential targets for
 PT agents against pathogenic bacteria.
 XX
 PS Claim 1; SEQ ID NO 79; 185pp; English.


```

PR 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX N-PSDB; ABD04540.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 19715; 455pp; English.
XX CC The invention relates to Pseudomonas aeruginosa polypeptides and the
XX CC polynucleotides encoding them. The sequences are useful in diagnosis and
XX CC therapy of pathological conditions, as molecular targets for diagnostics,
XX CC prophylaxis and treatment of pathological conditions resulting from a
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,
XX CC for the ability to bind a P. aeruginosa nucleic acid, as components of
XX CC effective antibacterial targets, as targets for antibacterial drugs,
XX CC including anti-P. aeruginosa drugs, as templates for recombinant
XX CC production of P. aeruginosa-derived peptides or polypeptides, as target
XX CC components for diagnosis and/or treatment of P. aeruginosa-caused
XX CC infection, and in detection of P. aeruginosa sequences or other sequences
XX CC of Pseudomonas species using biotech technology. Sequences ABO67826-
XX CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format from USPTO at
XX CC segdata.uspto.gov/sequence.html
XX CC Sequence 269 AA;

Query Match 11.9%; Score 307; DB 7; Length 269;
Best Local Similarity 21.6%; Pred. No. 1.1e-14;
Matches 108; Conservative 46; Mismatches 93; Indels 254; Gaps 11;

QY 1 MMSGLFIGATGKTHSTGLGTVSNINANTIGYKQOQVFDLPSQDLAIGSTGSGQPN 60
DB 9 MLSALVSKTGSADMLTISNNLANVSTGFKRDRAEFDQLLYQTRRPGQSTQDS 68
QY 61 QAGMGAQVGS-VRTIFTOGAFFPGNSVT-----DLAIGKGGFFQVLTLED-KVHYTRAGNF 113
DB 69 ELPSGLQGTGVVRVGTQKIFTPGSLQTTQEPDLMAVNGRGGFFQVLLPDGTVSVTRDGSF 128
QY 114 RFTQDGLNDPSGFTLMGSRISNNFNKKTLEPTQLDFNDPTVAKSPAKTSTALNAVN 173
DB 129 HUNSDQIVTSGFAL-----EPAL----- 148
QY 174 LGDSTDKTQSEANPYFALLSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGA 233
DB 149 ----- 148
QY 234 PSTGSKTFEYLIVANPSEDGSAASGTSAGLLMSGTMTFSSGELKMTAFTTGSATK 293
DB 149 -----VVPNE-----TQTF----- 158
QY 294 DLNMQAPLVLNGLPQFSGANFYVAGIQIPLTLDFGIKSQQNNWAGAPASAAIGTDIGKLP 353
DB 159 -----VG----- 160
QY 354 SMPQTSSGNTARGSSSTRYSQDGYPOGDLVDVITITSEGLKQKYSNQVDFYNI 413
DB 161 -----QDG-----TVSVTTGNAQ-----PQVIG--NI 181
QY 414 PLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTNSYKLSVNQLETSNVDMRSRMVNM 473
DB 182 QTADFINPAGLQAIQNNLPLETSSGSAQVGTGPGNLGLGTVAQNTLNSNVVVEELVNM 241
QY 474 IIIQGFQNNKSVTTADTML 494
DB 242 ITTORAYEMNSKVIQTADQML 262

RESULT 14
ABM68266
ID ABM68266 standard; protein; 261 AA.
XX AC AEM68266;
XX DT 20-NOV-2003 (first entry)
XX DE Photorhabdus luminescens protein sequence #1363.
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX KW whooping cough.
XX OS Photorhabdus luminescens.
XX PN WO200294867-A2.
XX PD 28-NOV-2002.
XX PF 07-FEB-2002; 2002WO-IB003040.
XX PR 07-FEB-2001; 2001FR-00001659.
XX PA (INSP ) INST PASTEUR.
XX PA (CNRS:) CNRS CENT NAT RECH SCI.
XX PI Duchaud E, Tacurit S, Glaser P, Frangeul L, Kunst P, Danchin A;
XX PI Buchrieser C;
XX DR WPI; 2003-148459/14.
XX CC Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX PS Claim 2; SEQ ID NO 1363; 1205pp; French.
XX CC The invention relates to the isolation of genes and their encoded
XX CC proteins from Photorhabdus luminescens. The isolated sequences are
XX CC sources of probes and primers for detecting the genome of P. luminescens
XX CC and related species; to study polymorphisms; for gene analysis and for
XX CC detection/amplification of the genes. Antibodies (Ab) raised against the
XX CC polypeptides encoded by the genes are used for detection/identification
XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX CC carry a gene-containing vector are used to select compounds that
XX CC modulate, regulate, induce or inhibit expression of the genes in plants,
XX CC animals or microorganisms other than P. luminescens and are able to alter
XX CC response or sensitivity to toxins and antibiotics produced by P.
XX CC luminescens. Cells transformed to express the genes are useful for
XX CC recombinant production of the proteins, particularly toxins and
XX CC antibacterials useful as insecticides, bactericides and fungicides. The
XX CC genes, proteins, vectors containing the genes and Ab are also useful
XX CC therapeutically (to treat microbial infection by bacteria or fungi that
XX CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX CC biopesticides. Other uses of the genes and the proteins are as virulence
XX CC factors and for identifying targets of human diseases for which P.
XX CC luminescens is a model (particularly plague and whooping cough). This
XX CC sequence represents one of the isolated P. luminescens proteins
XX SQ Sequence 261 AA;

Query Match 11.4%; Score 296; DB 6; Length 261;
Best Local Similarity 20.3%; Pred. No. 7.3e-14;
Matches 104; Conservative 53; Mismatches 93; Indels 262; Gaps 8;

QY 1 MMSGLFIGATGKTHSTGLGTVSNINANTIGYKQOQVFDLPSQDL-----AIGSTGS 56
DB 1 MRSLSWIATGDLDAQQTNDVIANNLANVSTGFKRQRAIFEDLLYQNIQPGANSSQT 60
QY 57 QQGN--QAGMGAQVGSVRTIFTOGAFFPGNSVTDIAIGKGGFFQVLTLEDKVH-YTRAGNF 113
```

CC	pylori infection or to identify H. pylori polypeptide binding compounds,				
CC	useful as potential H. pylori life cycle activators or inhibitors. The				
CC	genomic sequence of H. pylori (ATCC 55679) was determined from				
CC	overlapping contigs generated by mechanically shearing the bacterial DNA.				
CC	The sequences were analysed for ORF of at least 180 nucleotides, and the				
CC	predicted coding regions defined by computer evaluation. To identify				
CC	likely H. pylori antigens for vaccine development, the amino acid				
CC	sequences predicted from various ORF were analysed for significant				
CC	homology to other known or exported membrane proteins. Having identified				
CC	and determined the sequences of interest, particular regions can be				
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide				
CC	production, e.g. in E. coli hosts				
XX	Sequence 271 AA;				
QY	Query Match	11.1%;	Score 288;	DB 2; Length 271;	
DB	Best Local Similarity	21.0%;	Pred. No. 3.1e-13;		
QY	Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;				
QY	1	MMGSLFIGATGKMTHTGLGTGVNNIANANTIGYQQQVVFQDLFQSDLAIGTSGSQGN	60		
DB	10	MLRSLSYATSQMLAQQTHIDTTNNIANVNTGFKSRADFNDFYQAMQVAGTNTSNT	69		
QY	61	-----QAGMGAQVGSVRTFTQGAPEGNSVTDLAIGKGFQVTLFD-KVHYTRAGNF	113		
DB	70	LSPDGMVEGLGVRSPAITKMFQSGPKETENMLDIAITGKGFQVQLPDGTTAYTRSGNF	129		
QY	114	RFTQDGFINDPSPFTLMSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNN	173		
DB	130	KLDEQNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN	159		
QY	174	LGDSTDKTQSEANPYFALLESKWNGTTPPISTNSYSYAQPMRVYDQGNSHDITVYFDGA	233		
DB	160	IG-----VDGT	165		
QY	234	PSSTGSKTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK	293		
DB	166	VSVT-----	169		
QY	294	DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKSQNMWAGAPASAAAIGTDIGKLP	353		
DB	170	-----	169		
QY	354	SMWPIQTSSGNSTARNGSSSTRYSQDGYPQGLVDVTITSEGKLGKYSNSQVVDVYNI	413		
DB	170	-----QGLQTTNSVIG--QI	182		
QY	414	PLARFTSGDLRREGNNHYSATLDGGPEFGLPGTSNYGKLSVNOLETNSVDMRSRNVNM	473		
DB	183	TLANFVNPAGLHSMGDNLFSTINASGDAIVGNPDQSGGLKRLQCGFLELSNVLVEEMTDL	242		
QY	474	IIIQRFQFOMNSKSVTTADTMLOKALELKR	502		
DB	243	ITAQRAYEANSKSIQTADAMLQTVNSLKR	271		
RESULT 16					
AAW24645					
ID	AAW24645	standard; protein; 271 AA.			
XX	AAW24645;				
DT	11-AUG-1997 (first entry)				
XX	H. pylori flagella associated protein, 29298130.aa.				
XX	Chronic gastritis; duodenal ulcer disease; activator; inhibitor;				
KW	bacterial life cycle; vaccine; immunisation; detection; antisense;				
KW	inhibition; flagella; flagellum; basal body; rod.				
OS	Helicobacter pylori.				
PN	WO9719098-A1.				

DB	61	RLPSGLIGTVRVAERHISQNLSETNSHNVAIRGKGFQVQLPDGTSAYTRDGSF	120		
QY	114	RFTQDGFINDPSPFTLMSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNN	173		
DB	121	QEDQNGQLTTANGELIF-----PFI-----	140		
QY	174	LGDSTDKTQSEANPYFALLESKWNGTTPPISTNSYSYAQPMRVYDQGNSHDITVYFDGA	233		
DB	141	-----	140		
QY	234	PSSTGSKTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK	293		
DB	141	-----	140		
QY	294	DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKSQNMWAGAPASAAAIGTDIGKLP	353		
DB	141	-----	140		
QY	354	SMWPIQTSSGNSTARNGSSSTRYSQDGYPQGLVDVTITSEG-----KLQKYSNSQVVD	409		
DB	141	-----TIPDNAT-----DLTISDGIYSVKVQGSVPQVQVG	172		
QY	410	MYNIPLARFTSGDLRREGNNHYSATLDGGPEFGLPGTSNYGKLSVNOLETNSVDMRSR	469		
DB	173	F---TLTTFINDSGLESIGENLYVETSSGVPIENTPGINGAGLLQHKYVETSNVIAEE	229		
QY	470	MVNMIIQRFQFOMNSKSVTTADTMLOKALELK	501		
DB	230	LVMNIQTRAYEVNSKAVSDQMLQKLTQLZ	261		
RESULT 15					
AAW20373					
ID	AAW20373	standard; protein; 271 AA.			
XX	AAW20373;				
DT	29-JUL-1997 (first entry)				
XX	H. pylori flagella associated protein, 29298130.aa.				
XX	Vaccine; prevention; treatment; infection; identification;				
KW	binding compound; bacterium; life cycle; activator; bacteria;				
KW	duodenal ulcer disease; chronic gastritis; diagnosis; flagella;				
KW	associated; flagellum.				
XX	Helicobacter pylori.				
OS	WO9640893-A1.				
FN	19-DEC-1996.				
XX	06-JUN-1996; 96WO-US009122.				
XX	07-JUN-1995; 95US-00487032.				
PR	01-APR-1996; 96US-00630405.				
XX	(ASTR) ASTRA AB.				
PI	Smith D, Berglinth OT, Mellgaard BL;				
XX	WPI; 1997-052306/05.				
DR	N-PSDB; AAT67783.				
XX	Helicobacter pylori nucleic acid sequences and related polypeptide(s) -				
PT	useful for vaccines to treat or prevent H. pylori infection, and to				
PT	detect Helicobacter.				
XX	Claim 56; Page 560-561; 1481pp; English.				
PS	The present sequence is a Helicobacter pylori flagella associated				
CC	protein. The protein may be used in a vaccine to prevent or treat H.				

XX PD 29-MAY-1997.
 XX XX 15-NOV-1996; 96WO-US018542.
 XX PF 17-NOV-1995; 95US-00561469.
 XX PR (ASTR) ASTRA AB.
 XX PA Smith DH;
 XX PI WPI; 1997-298052/27.
 XX DR N-PSDB; AAT77463.
 XX XX
 XX PT Helicobacter pylori nucleic acid sequences and related proteins - used
 XX PT for diagnostics and therapeutics.
 XX PS Claim 18; Page 167-168; 235pp; English.
 XX XX
 CC The present sequence is a Helicobacter pylori flagella associated
 CC protein, which was found to be homologous to flagellar basal body rod
 CC protein following BLAST protein analysis. H. pylori has been strongly
 CC linked to chronic gastritis and duodenal ulcer disease. The nucleic acid
 CC sequences of the invention are used to evaluate compounds, especially
 CC activators or inhibitors of bacterial life cycle, for the ability to bind
 CC an H. pylori nucleic acid sequence. The nucleic acid sequences, and
 CC corresponding proteins, are also useful for generating vaccines for
 CC immunising subjects against H. pylori or for use in detecting the
 CC presence of Helicobacter species in a sample. Antisense nucleic acid
 CC sequences of these sequences are used to inhibit expression of a gene
 CC from Helicobacter species. H. pylori whole genomic DNA was isolated and
 CC nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-
 CC ended and ligated to unique BstXI-linker adapters in 100-1000 fold molar
 CC excess. These linkers are complementary to the BstXI-cut pMPX vectors,
 CC while the overhang is not self-complementary. Therefore the linkers will
 CC not concatamerise nor will the cut vector re-ligate itself easily. The
 CC linker-adaptor inserts were ligated to each of the 20 pMPX vectors to
 CC construct a series of shotgun subclone libraries. The purified DNA
 CC samples were then sequenced. Note: The ORF/protein reference number for
 CC this sequence was obtained from the related specification, WO9640893
 XX
 XX Sequence 271 AA;

Query Match 11.1%; Score 288; DB 2; Length 271;
 Best Local Similarity 21.0%; Pred. No. 3, 1e-13;
 Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQQQVVFQDLPSQDLAIGTSGQGN 60
 DB 10 MLRSLSYATSGMLAQOQTHIDTSSNNIANVNTGFKSRADFNDFYQAMQVAGTNTNT 69
 QY 61 -----QAGMGAQGVSVRTITQGAFFPCNSVTDLAIGGKGFVOTLED-KVHYTRAGNF 113
 DB 70 LSPDGMVEVLGVPRPAITWMSQSPKTEENLDAITGKGFQVQVQPDGTTATRGNF 129
 QY 114 RPTQDGLNDPSGFTLMGSRISNNPNIKETLEFIQDLDFNDPTVAKSPAKTSTALNAV 173
 DB 130 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 159
 QY 174 LGDSYDTKTQSEANPYFALLESKNGKNGTPPISTNSVYAQPMRVYDQGNSHDIIVPDGA 233
 DB 160 IG-----VDGT 165
 QY 234 PSSTGSKTFEYLVAMNPSEDSAASTGDSAGLMSGTMTFSSNGELKNMTAFTPTGSATK 293
 DB 166 VSVT----- 169
 QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDLFGIKSQQNWAGAPASAAAGTIDGKLP 353
 DB 170 ----- 169
 QY 354 SMMPIQTSSGNSTARGSSSTRYSQDGYPOGLVDVTTITSEGKLGKYSNSQVDFYNI 413

DB 170 -----QGLQTSNVIG--OI 182
 QY 414 PLARFTSEDLRREGNNHYSATLDSGPGFLPGTSSNYGKLSVNOLETNSVDMKSEVMNM 473
 DB 183 TLANFNPAGLHSMGDNLFSTWASDAIVGNPDSQGLKRGQFLELSNVRLVEMTDL 242
 QY 474 IITQRFQMNKSVTTADTMLQKALELKR 502
 DB 243 ITAQRAYEANSKSIQTADAMLTQVNSLKR 271

RESULT 17
 AAW20863
 ID AAW20863 standard; protein; 279 AA.
 XX AC AAW20863;
 XX DT 16-JUL-1997 (first entry)
 XX DE H. pylori flagella-associated membrane protein, 12ge20305orf11.
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; flagella.
 XX OS Helicobacter pylori.
 XX PN WO9640893-A1.
 XX PD 19-DEC-1996.
 XX PF 06-JUN-1996; 96WO-US009122.
 XX PR 07-JUN-1995; 95US-00487032.
 XX PR 01-APR-1996; 96US-00630405.
 XX PA (ASTR) ASTRA AB.
 XX PI Smith D, Berglindeh OT, Mellgaard BL;
 XX DR WPI; 1997-052306/05.
 XX DR N-PSDB; AAT68116.
 XX PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 XX PS Claim 56; Page 1266-1267; 1481pp; English.
 CC The present sequence is a Helicobacter pylori flagella-associated
 CC membrane protein. The protein may be used in a vaccine to prevent or
 CC treat H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was determined
 CC from overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides, and
 CC the predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts
 XX SQ Sequence 279 AA;

Query Match 11.1%; Score 288; DB 2; Length 279;
 Best Local Similarity 21.0%; Pred. No. 3, 2e-13;
 Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQQQVVFQDLPSQDLAIGTSGQGN 60
 DB 18 MLRSLSYATSGMLAQOQTHIDTSSNNIANVNTGFKSRADFNDFYQAMQVAGTNTNT 77

QY	61	-----QAGMAGVGSVRTIFQGAPEGNSVTDLAIGGKGFQVLTLED-KVHYTRAGNF	113
Db	78	LSPDGEVLGVRPSAITKMFSGSKETENNLDIAITGKGFQVQLPDGTTAYTSGNF	137
QY	114	RFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVVN	173
Db	138	KLDEQNLVTSEGYLLI-----PQI---TL-----PEDTQO-----VN	167
QY	174	LGDSDTKTQSEANPYFALLESKMGNGTTPPISSTNSYSYAQPMRVYDQGNHSDITVYDGA	233
Db	168	IG-----VDGT	173
QY	234	PSSTGSKTFEYLVAMNPSEDGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK	293
Db	174	VSVT-----	177
QY	294	DLNAWQAPLVNGLPQFSANFVAGIQPLTLDFGIKSQQNMWAGAPASAAAIGTDIGKLP	353
Db	178	-----	177
QY	354	SNMPIQTSSGNSTARNGSSSTRYSQDGYPPQGLVDVTTITSEKLGKYSNQVDFYNI	413
Db	178	-----QGLQTTSNVIG--QI	190
QY	414	PLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDMSEWVM	473
Db	191	TLANFVNPAGLHSGNDLFSITNASGDIAIVGNPDQGLGKLRQGFLELSNVRIVEEMTDL	250
QY	474	IITQGFQMNKSVTVTADTMLOKALELKR	502
Db	251	ITQRAYEANSKSIQTADAMLOTVNSLKR	279
RESULT 18			
ID	ADF07825	standard; protein; 265 AA.	
XX	AC	ADF07825;	
XX	DT	12-FEB-2004 (first entry)	
XX	DE	Bacterial polypeptide #3938.	
XX	KW	Proteus mirabilis infection; bacterial infection; antibacterial;	
XX	KW	immunostimulant.	
XX	OS	Proteus mirabilis.	
XX	PN	US6605709-B1.	
XX	PD	12-AUG-2003.	
XX	PF	05-APR-2000; 2000US-00543681.	
XX	PR	09-APR-1999; 99US-0128706P.	
XX	PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX	PI	Breton GL;	
XX	DR	WPI; 2003-895291/82.	
XX	DR	N-PSDB; ADF03653.	
XX	PT	New Proteus mirabilis polypeptides and polynucleotides, useful as	
XX	PT	reagents for diagnosis of bacterial disease, as components of	
XX	PT	antibacterial vaccines, as targets for antibacterial drugs, or as	
XX	PT	biocontrol agents for plants.	
XX	PS	Disclosure; SEQ ID NO 8110; 870pp; English.	
XX	XX	The invention relates to new Proteus mirabilis polypeptides and	
CC	CC	polynucleotides. The invention also relates to antibodies against the	

CC	polypeptides, methods for producing the polypeptides, a method of		
CC	generating vaccines for immunising an individual against P. mirabilis, a		
CC	method for evaluating a compound for the ability to bind a P. mirabilis		
CC	polypeptide and a method for screening test compounds for anti-bacterial		
CC	activity. The polypeptides and polynucleotides are useful as molecular		
CC	targets for diagnosing, preventing and treating pathological conditions		
CC	resulting from bacterial infection, as reagents for diagnosis of		
CC	bacterial diseases, as components of antibacterial vaccines, as targets		
CC	for antibacterial drugs or as bio-control agents for plants. This		
CC	sequence represents a Proteus mirabilis polypeptide of the invention.		
XX	Sequence 265 AA;		
QY	Query Match		
Db	Best Local Similarity		
QY	Matches 100; Conservative		
Db	Score 286; DB 7; Length 265;		
QY	11.1%; Pred. No. 4.2e-13;		
Db	55; Mismatches 98; Indels 254; Gaps 9;		
QY	1 MWGSLFTGATGKTHSTGLGTVSNNIANANTIGYKQGVQDFLFSQDL-----AIGTGS		
Db	6 MIRSLLIAKTGLDAQQTNDMDVISNNLANVSTNGFRQRAVDFDLYLQTIROFGAMTSEQT		
QY	57 QGPN--QAGMAGVGSVRTIFQGAPEGNSVTDLAIGGKGFQVLTLEDKVH-YTRAGNF		
Db	66 NAPSGLQIGTVRPVATERLHSHSQGNLAQTNTNRDVALKGGFFHVQLPDGTDATRDGSF		
QY	114 RFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVVN		
Db	126 QMDQNGQLVTSGGFQIVPA-----IILPETAKKVMV-----		
QY	174 LGDSDTKTQSEANPYFALLESKMGNGTTPPISSTNSYSYAQPMRVYDQGNHSDITVYDGA		
Db	157 -----ORDGIVSVEIEGS		
QY	234 PSSTGSKTFEYLVAMNPSEDGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK		
Db	170 PA-----		
QY	294 DLNAWQAPLVNGLPQFSANFVAGIQPLTLDFGIKSQQNMWAGAPASAAAIGTDIGKLP		
Db	172 -----PQ-----QVQQL-178		
QY	354 SNMPIQTSSGNSTARNGSSSTRYSQDGYPPQGLVDVTTITSEKLGKYSNQVDFYNI		
Db	179 -----TLTT-----		
QY	414 PLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDMSEWVM		
Db	183 -----FINDSGLESVGENLYLETASSCAPTENAPNGINGAGLLYQGYVETSNVNAEELVNM		
QY	474 IITQGFQMNKSVTVTADTMLOKALEL		
Db	239 IOTQRAYEANSKSIQADAMLOKLTQL		
RESULT 19			
ID	AA51377	standard; protein; 263 AA.	
XX	AC	AA51377;	
XX	DT	04-MAY-2000 (first entry)	
XX	DE	C. jejuni flag protein.	
XX	KW	Basal body rod protein; flag; flagellum; vaccine; immunogen;	
XX	KW	pathogenic bacteria; detection; antibacterial.	
XX	OS	Campylobacter jejuni.	
XX	PN	US6020125-A.	
XX	PD	01-FEB-2000.	
XX	XX		

PF 07-JUN-1995; 95US-00483857.
 PR 08-MAY-1995; 95US-00436748.
 XX (CONN-) CONNAUGHT LAB LTD.
 PA Louie H, Chan VL;
 PI WPI; 2000-146875/13.
 DR N-PSDB; AA288554.
 XX Recombinant basal body rod protein producible by a transformed host for
 PT use in immunological vaccine compositions for in vivo administration to
 PT protect against diseases caused by bacterial pathogens.
 XX
 XX Example 1; Col 23-26; 28pp; English.
 XX This invention describes a novel recombinant basal body rod protein (I)
 CC producible by a transformed host containing an expression vector
 CC comprising a nucleic acid selected from; (a) an entire nucleic acid
 CC sequence of 1800 base pairs (bp), or the nucleic acid sequence of the
 CC FlgG gene having 810 bp, both given in the specification; (b) a nucleic
 CC acid encoding the amino acid sequence of the flgG protein, a 270 residue
 CC sequence, given in the specification; (c) a nucleic acid sequence
 CC encoding a functional flgG basal body rod protein of a flagellum of a
 CC strain of Campylobacter; or (d) an immunogenic fragment of an flgG
 CC protein of (a), (b) or (c), and expression means operatively coupled to
 CC the nucleic acid molecule for expression by the host of a basal body rod
 CC protein of a flagellum of a strain of Campylobacter. (I) is useful in
 CC immunological vaccine compositions for in vivo administration to protect
 CC against diseases caused by bacterial pathogens that produce basal body
 CC rod proteins. (I) are also useful as immunogens, as antigens in
 CC immunoassays, or for procedures for the detection of antibacterial,
 CC Campylobacter, basal body rod protein and/or peptide antibodies. The
 CC product of the invention has immunostimulatory activity. This sequence
 CC represents the Campylobacter jejuni flgG protein described in the method
 CC of the invention
 XX Sequence 263 AA;
 XX
 Query Match 9.2%; Score 237.5; DB 3; Length 263;
 Best Local Similarity 18.6%; Pred. No. 1.8e-09;
 Matches 95; Conservative 45; Mismatches 115; Indels 255; Gaps 9;
 QY 1 MMSLFIGATGCMKTHSTGLGTVSNINANTIGYKQOQVVDLFSQDLAIGTSGQGN 60
 DB 1 MMSLHTAATGVAQQQIDVTSNNIANVNTAGPKSRABFADLMYQVMKYAGTSTATT 60
 QY 61 QAGMGAQVG-----SVRTIFTOGAFE-PGNSVTDLAIGKGFQVLTLED-KVHYTRAGN 112
 DB 61 LSPSGIEVGVRPTAVTKVFTGKSLKSTSDGLDMAIGAGFFQIQLPDGTIGYTRNGQ 120
 QY 113 FRFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172
 DB 121 FTKNEGVNVDGYRL----- 137
 QY 173 NLGSDTDKTOSEANPYFALLESMKNGTGPPISTSNYSYAQPMRVYDQGNSHDITVYFDG 232
 DB 138 ----- 137
 QY 233 APSSTGSKTEYLVMNPSEGGASCTDSAGLLMGSTMTFTSSNGELKNWTAFTPTGSAT 292
 DB 138 ----- 137
 QY 293 KDLNMQPAPLVNGLPOFSANFVAGIQPLTLDGFKSQQNMWAGAPASAAIGTDIGKL 352
 DB 138 -----LPE-----WTIEG-----ATAINVATD-GRV 158
 QY 353 PSMPIOTSSGNSTARNSSSTRYSQDGYPOGDLVDVITITSEKLGQKYSNQVDFYN 412
 DB 159 SVMLP-----GEQQTQI-----GQ 173
 QY 413 IPLARFTSEDGLRREGNNHYSATLDSGGPFGLPGTSNYKLSVNLQETSNVDMSREYN 472

DB 174 VELVQFINPAGLHSMGDNLYLETCASGAPVAGIAGQDGLGTRHGFIELSNVQLEMTD 233
 QY 473 MIIIRGFQFMNSKSVTTADTMLQKALELKR 502
 DB 234 LITQRAYEAGSKAITTSDMLGIVNQLKK 263
 RESULT 20
 AAW73073
 ID AAW73073 standard; protein; 263 AA.
 XX
 AC AAW73073;
 XX
 DT 05-JAN-1999 (first entry)
 XX
 DE FlgG protein.
 XX
 KW FlgG operon; flgG protein; flgG protein; basal body rod protein;
 KW flagellum; secretory diarrhoea; enteritis; vaccine; therapy.
 XX
 OS Campylobacter jejuni.
 XX
 PN US5827654-A.
 XX
 PD 27-OCT-1998.
 XX
 PF 08-MAY-1995; 95US-00436748.
 XX
 PR 08-MAY-1995; 95US-00436748.
 XX
 PA (UTOR) UNIV TORONTO.
 XX
 PI Louie H, Chan VL;
 XX
 DR WPI; 1998-593983/50.
 DR N-PSDB; AAV58977.
 XX
 PT DNA encoding Campylobacter flagellum basal body rod proteins - useful for
 PT recombinant production of the proteins for use as vaccines against the
 PT bacterium, and for its detection, additionally with antibodies raised
 PT with the protein.
 XX
 PS Claim 1; Fig 1; 27pp; English.
 XX
 CC This sequence represents the Campylobacter flgG protein, encoded by the
 CC flgG operon of the invention. The flgG and flgG proteins are basal body
 CC rod proteins of the flagellum of the Campylobacter. The flgG operon and
 CC host cells containing it, are useful for the recombinant production of
 CC Campylobacter, especially C. jejuni, basal body rod proteins. This
 CC bacterium is the cause of secretory diarrhoea and enteritis. As such, the
 CC peptides produced can be used to raise antibodies, which in turn can be
 CC used to detect the presence of the organism diagnosis of the conditions.
 CC Additionally, the peptides, and specifically the live vectors (e.g. pox-
 CC or vaccinia virus) can be used as vaccines against the bacterium, and the
 CC antibodies can be used for passive immunisation. The nucleic acids can
 CC also be used to detect the presence of the bacterium
 XX
 SQ Sequence 263 AA;
 Query Match 8.6%; Score 223.5; DB 2; Length 263;
 Best Local Similarity 19.3%; Pred. No. 2e-08;
 Matches 98; Conservative 43; Mismatches 115; Indels 253; Gaps 12;
 QY 1 MMSLFIGATGCMKTHSTGLGTVSNINANTIGYKQOQVVDLFSQDLAIGTSGQGN 60
 DB 1 MMSLHTAATGVAQQQIDVTSNNIANVNTAGPKSRABFADLMYQVMKYAGTSTATT 60
 QY 61 QAGMGAQVG-----SVRTIFTOGAFE-PGNSVTDLAIGKGFQVLTLED-KVHYTRAGN 113
 DB 61 LSPSGIEVGVRPTAVTKVFTGKSLKSTSDGLDMAIGAGFFQIQLPDGTIGYTRNGQ 120
 QY 113 FRFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDNDFTVAKSPAKTSTALNAV 173

109 -----PDGTIEMG-----NLQK-----IMK----- 123
174 LGDSTDKTQSEANPYFALLEWSKNGTPISTNSYVAQPMRVYDOQGNSHDITVYFDGA 233
124 -----DNEGNI----- 129
234 PSSTGSKTFEYLVAMNPSEDGSAASGSDSAGLIMSGTMTFSSNGELKNMTAFTPTGSATK 293
130 -----VNSDGYRLLPETI-----PEG----- 146
294 DLNAWQAPLVNGLPOFSANFVAGIQPLTLDGIXSQNMWAGAPASAAAIGTDIGKLP 353
147 -----ATAINVATD-GTYS 159
354 SMMPIOTSSGNSTARNGSSSTRYSQGYPOGDLVDVITITSEKLOGKYSNSQVDFYNI 413
160 VMLP-----GEQEQEI-----GOV 174
414 PLARFTSEDLREGNNHYSATLDSGPEPGLPGTNYGKLSVNQLETNSVDMRSWVM 473
175 ELVQFINPAGLHSMGDNLYLETGASGAPVAGIAGODGLGTIRHGFIELSNVQLVEEMTDL 234
474 IITORGQWMSKSVTTADTMQLKALELKR 502
235 ITQRAYEAGSKAITTSDDDMLGIVNQLKR 263

RESULT 21
ABO81481
ID ABO81481 standard; protein; 2736 AA.
XX AC ABO81481;
XX XX
29-JUL-2004 (first entry)
Pseudomonas aeruginosa polypeptide #13656.
XX XX
Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX XX
Pseudomonas aeruginosa.
XX OS
US6551795-B1.
XX FN
22-APR-2003.
XX PD
18-FEB-1999; 99US-00252991.
XX PF
18-FEB-1998; 98US-0074788P.
XX PR
27-JUL-1998; 98US-0094190P.
XX PR
(GENO-) GENOME THERAPEUTICS CORP.
XX PA
Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX FI
WPI; 2003-615309/58.
XX DR
N-PSDB; ABD15052.
XX DR

Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from bacterial infection.
XX PT
XX PT
XX XX
XX PS
Disclosure; SEQ ID NO 30227; 455pp; English.
XX CC
The invention relates to Pseudomonas aeruginosa polypeptides and the polynucleotides encoding them. The sequences are useful in diagnosis and therapy of pathological conditions, as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from a bacterial infection, for evaluating a compound, such as a polypeptide, for the ability to bind a P. aeruginosa nucleic acid, as components of effective antibacterial targets, as targets for antibacterial drugs, including anti-P. aeruginosa drugs, as templates for recombinant production of P. aeruginosa-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of P. aeruginosa-caused infection, and in detection of P. aeruginosa sequences or other sequences of Pseudomonas species using biochip technology. Sequences ABO81481-ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html
XX XX
SQ Sequence 2736 AA;

Query Match 6.8%; Score 176; DB 7; Length 2736;
Best Local Similarity 22.6%; Pred. No. 0.0021;
Matches 120; Conservative 56; Mismatches 212; Indels 142; Gaps 22;

QY 20 GTVNNNIANTIGYKQOQVFDLPQDLAIGSGTSGOG-----PNOAGYGAQGVSV 71
DB 1320 GTVVNAVA-----QDPA-GNTGPGGTTVDVAPNTPVVPNSGNL 1359
QY 72 RTIFTQGAFFPGNSVTDLAIGGKGFQVTLKDKVHYTRAGNFRFTDGLNDPSGFTLMG 131
DB 1360 -----LNGTAEGSVTLTDGNGNPIGQTADG-----SGNWSFTPGSQLPNTVAVNTA 1409
QY 132 SRISNNPNIKKETLEPIOLDENDPTVAKSPAKTALNAVNLGSDTDKTOSEANPYFAL 191
DB 1410 SDAAGNTSAPATT-----TVDSLSFIPQVDPSPNGSVISGTADAGNTIITDNGNPIGQV 1465
QY 192 LESWKGNG--TPPISTSNYSVAQPMRVYDOQGNSHDITVYFDG-APSGTSGKTFEYLVAM 248
DB 1466 TADSGNWSFTPGPLPDTGVNVVARSPPSVNDSAPAVITVDGYAPAP-----VI 1516
QY 249 NPSEDGSAASGTDGAGLLM-----SGTMTFS-----SNGELKNMTAFT 286
DB 1517 DPS-NGTEISGTAEAGATVILTDGNGNPIGQATADGSGNWTFTPTPLANGTINAVAO 1575
QY 287 PTG-----SATKDLNANQAPLVN-----GLPQPSANFV-----GAGIQPLTLD 327
DB 1576 PAGNTSGPASVTVDIAA-PPAPVNPNSGNGVVISGTAEAGATVILTDGNGNPIGQVTD 1634
QY 328 IKSQNMWAGAPASAAAIGTDIGKLPNMP:QTSNGNSTARNGSSSTRYSQGYPOGDL 387
DB 1635 GK-----WAFIPATPLANGTVINAL-----AQDAAGNNSPTSATVDSLAPAAPV 1679
QY 388 VDVTTTSEKLOGKYSNSQVDFYNIPLARTSEDLREGNNHYSATLDSGPEPGLPG 447
DB 1680 IDP-----SNGSVIAGTAEAGATVILTDG-----NGNPIGQVTDGSGNWSFTPG 1724
QY 448 TSNYKGLSVNQLETNSVDMRSWVMNIIQRFQVNSKSVTTADTMLOKA 497
DB 1725 T-----PLSNGTV-----VNAVAQDRAAGTSGPASTTVDVSVAPAA 1759

RESULT 22
AAB01847
ID AAB01847 standard; protein; 1095 AA.
XX AC AAB01847;
XX XX
12-SEP-2003 (revised)
DT 11-SEP-2000 (first entry)
XX XX
Haemophilus influenzae strain 12 mature HMWIA protein, SEQ ID NO:69.
XX XX
Mature HMW protein; hmw gene; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX XX
Haemophilus influenzae; strain 12.
OS
WO2000020609-A2.
XX PN
13-APR-2000.
XX PD

XX PF 07-OCT-1999; 99WO-CA0000938.
 XX PR 07-OCT-1998; 98US-00167568.
 XX PR 08-DEC-1998; 98US-00206942.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Loosmore SM, Yang Y, Klein MH;
 XX DR WPI: 2000-303789/26.
 XX DR N-PSDB; AA52196.
 XX PT Nucleic acid molecule for producing recombinant high molecular weight
 XX PT proteins of Haemophilus which are used as a vaccine to provide protection
 XX PT against Haemophilus induced diseases in humans.
 XX PS Claim 8; Fig 28F-Q; 307pp; English.
 XX CC The invention relates to the recombinant production of Haemophilus
 XX CC influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The
 XX CC expression construct used to effect recombinant expression comprises a
 XX CC promoter functional in *E. coli* (e.g., the T7 promoter) operably linked to
 XX CC a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 XX CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 XX CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 XX CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
 XX CC and the hmwB and hmwC genes encode accessory proteins which are
 XX CC responsible for post-translational processing and secretion of the HMWA
 XX CC proteins. The modified hmwABC operon used in the expression construct of
 XX CC the invention contains an A gene modified such that it encodes only the
 XX CC mature HMWA. The invention also discloses hmwA genes (AA52175-AS2198)
 XX CC and HMWA proteins (AA501824-801849) from the non-typeable H. influenzae
 XX CC strains J09C, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 XX CC vectors are used for the production of recombinant H. influenzae HMW
 XX CC proteins which can be used as vaccines to mediate a humoral or cell-
 XX CC mediated immune response to provide protection against diseases in humans
 XX CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and
 XX CC tracheobronchitis). The HMW proteins are also useful as antigens in
 XX CC immunoassays for detecting antibodies against Haemophilus, HMW proteins
 XX CC and/or HMW peptides. The nucleotide sequences encoding the HMW proteins
 XX CC can be used to isolate and clone hmw genes from other non-typeable
 XX CC strains of Haemophilus via hybridisation reactions. The present sequence
 XX CC represents a mature HMWA protein from a non-typeable strain of H.
 XX CC influenzae. (Updated on 12-SEP-2003 to standardise OS field)
 XX SQ Sequence 1095 AA;

Query Match 6.7%; Score 173.5; DB 3; Length 1095;
 Best Local Similarity 22.8%; Pred. No. 0.00088;
 Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKHTSTGLGTYSNNIANAN-TIGYKQOVVFQDLFSODLAIGTSGQPNQAGMGQVGS 70
 DB 570 IKAGVDGNSDSTANNANLTIKTKELK-----TQDLNI-----SGFNKAEITAKDGS 618
 QY 71 VRTFTQGAFFPGNS-----VTDLAIGGKGFQVLTEDKHVYTRAGNFRFTQDGFND 123
 DB 619 DLTIGNTNSADGTNAKVTNFQVSKISADG-HKVTLHRSKVE--TSGSNNTEDSSDNN 675
 QY 124 PSGETMGSRISNPNKETEPIQLDFNDPTVAKSPAKTSTALNANVNLGDSPTKQGS 183
 DB 676 -AGLTIDAKNVTNNNTTSHKAVSI-----SATSGEITTKTGTINATTGNVEITATQGS 729
 QY 184 EANFYFALLSRSKNGH-----PPISNTSYSAQPMRVYDQGNSHDITVYFPGAPSSTG 238
 DB 730 -----ILGGIESGGSVLTATEGALVSNIS-----GNTVTVTAN-SCALTTLA 773
 QY 239 SKTFEYLVANPNPDEGSAAGTDSAGLL---MSGTWTFSNGELKNWTFPTFGSATKDL 295
 DB 774 GSTIKGTESVTTSSQSDIGTISGGTVEVKATESLTTQSNKIKATTGEANVTSATGTI 833
 QY 296 NAWQPAFLVNLGPQFSANFVGAGTQPLTLDGFKISQQNMWAGAPASAAAIIGTIDGKLP 355

DB 834 GGTISGNTVN-----VTANAGDLTVNGAE-----INATECAATLTSSGKL----- 875
 QY 356 MPIQTSSGNSSTARNGSSSTRYSQDGYPOGDL--VDVTTTSEGKL---QKG----- 401
 DB 876 ----TTEASSHITSAGQVNLGAQDGSVAGSINAANVTILNTTGTTLTKVGSNNATSGTL 931
 QY 402 YNSQVVDYFNIPLARFTSEDLRREGNNHYATLDSGCPGFLPCTSNYGKLSVNLQLET 461
 DB 932 VINAKDAELNGAALGNHTVTVNATNANGSGSVIATSS---RVNITG----DLITINGL-- 982
 QY 462 SNVDMRSREMNMMIIITQGFQMNKSVTTADTNLQKALELKR 502
 DB 983 -NI-TSKNGINTVLL-KGVKIDVKYIQGIASVDEVEAKR 1020

RESULT 23
 AAR41723
 ID AAR41723 standard; protein; 1536 AA.
 XX AC AAR41723;
 XX DT 25-MAR-2003 (revised)
 XX DT 26-APR-1994 (first entry)
 XX DE High molecular weight protein 1 (HMW1).
 XX KW HMW; high molecular weight protein; virus; vaccine; influenza; epitope;
 XX KW immunity; haemophilus influenzae.
 XX OS Haemophilus influenzae.
 XX PN W09319090-A1.
 XX PD 30-SEP-1993.
 XX PF 16-MAR-1993; 93WO-US002166.
 XX PR 16-MAR-1992; 92GB-00005704.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Barenkamp SJ;
 XX DR WPI: 1993-320683/40.
 XX DR N-PSDB; AAQ49506.
 XX PT High molecular weight surface proteins - of non-typeable haemophilus
 XX PT which exhibit immunogenic properties.
 XX PS Claim 3; Fig 2; 100pp; English.
 XX CC The isolation and purification of the high molecular weight protein
 XX CC enables the identification of the major protective epitopes of the
 XX CC protein by conventional epitope mapping. These epitopes can then be
 XX CC synthesised using standard techniques and incorporated into fully
 XX CC synthetic or recombinant vaccines. (Updated on 25-MAR-2003 to correct PN
 XX CC field.)
 XX SQ Sequence 1536 AA;

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
 Best Local Similarity 22.8%; Pred. No. 0.0014;
 Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKHTSTGLGTYSNNIANAN-TIGYKQOVVFQDLFSODLAIGTSGQPNQAGMGQVGS 70
 DB 1011 IKAGVDGNSDSTANNANLTIKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059
 QY 71 VRTFTQGAFFPGNS-----VTDLAIGGKGFQVLTEDKHVYTRAGNFRFTQDGFND 123
 DB 1060 DLTIGNTNSADGTNAKVTNFQVSKISADG-HKVTLHRSKVE--TSGSNNTEDSSDNN 1116

QY 124 PSQFTLMGSRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTOS 183
DB 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGETTKTGTINATGNVEITAQGS 1170
QY 184 EANPYFALLESMKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGT 238
DB 1171 ----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTITLA 1214
QY 239 SKTFEYLVANMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTTGSATKDL 295
DB 1215 GSTIKGESVITSSQSGDIGTISGGTVEVKATESLTTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPMS 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQDGL---VDVTITSEGL---QK----- 401
DB 1317 ----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTVKGSNNINATSGTL 1372
QY 402 YSNQVDFYNIPLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTS---RVNITG---DLITINGL-- 1423
QY 462 SNVDMSEVMNMIIOQGFQNMKSQSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKXGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 24
ID AAR41725
XX AC AAR41725;
XX DT 25-MAR-2003 (revised)
XX DT 26-APR-1994 (first entry)
XX DE High molecular weight protein 1 (Hmw1).
XX KW Hmw; high molecular weight protein; virus; vaccine; influenza; epitope;
XX KW immunity; haemophilus influenzae; gene cluster.
XX OS Haemophilus influenzae.
XX FH Key Location/Qualifiers
XX FT Misc-difference 668..677
XX FT /note= "Possibly incorrect sequence. Alternative sequence
XX FT for this region is LNVSEGEFN. (See comments)."

XX FN WO9319090-A1.
XX PD 30-SEP-1993.
XX PF 16-MAR-1993; 93WO-US002166.
XX PR 16-MAR-1992; 92GB-00005704.
XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Barenkamp SJ;
XX DR WPI; 1993-320683/40.
XX DR N-PSDB; AAQ49508.
XX High molecular weight surface proteins - of non-typeable haemophilus
XX PT which exhibit immunogenic properties.
XX PS Claim 3; Figure 2/10; 100pp; English.
XX The isolation and purification of the high molecular weight protein
XX CC enables the identification of the major protective epitopes of the
XX CC protein by conventional epitope mapping. These epitopes can then be

CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines. This sequence is claimed to be the
CC same as that given in AAR41723 (High molecular weight protein 1) although
CC it does differ slightly. (Repeated regions which are possibly incorrect
CC and occur in the corresponding nucleotide coding sequence contribute to
CC these differences). (Updated on 25-MAR-2003 to correct FN field.)
XX SQ Sequence 1536 AA;

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.0014;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTSTGLGTYSNNIANAN-TIGYKQOQVFDLFSQDLAIGTSGSQPNQAGGAQVGS 70
DB 1011 IKAGVDGSDSDATNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTFTQCAFPGNS-----VTDLAIGGKFFQVLEDKVHVTRAGNPRFTQDGLND 123
DB 1060 DLTIGNTNSADGTNAKVTFFNQVKDSLISADG-HKVTILHSKE--TSGSNNNTEDSDNN 1116
QY 124 PSQFTLMGSRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTOS 183
DB 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGETTKTGTINATGNVEITAQGS 1170
QY 184 EANPYFALLESMKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGT 238
DB 1171 ----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTITLA 1214
QY 239 SKTFEYLVANMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTTGSATKDL 295
DB 1215 GSTIKGESVITSSQSGDIGTISGGTVEVKATESLTTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPMS 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQDGL---VDVTITSEGL---QK----- 401
DB 1317 ----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTVKGSNNINATSGTL 1372
QY 402 YSNQVDFYNIPLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTS---RVNITG---DLITINGL-- 1423
QY 462 SNVDMSEVMNMIIOQGFQNMKSQSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKXGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 25
ID AAR63505
XX AC AAR63505 standard; protein; 1536 AA.
XX AC AAR63505;
XX DT 25-MAR-2003 (revised)
XX DT 25-JUN-1995 (first entry)
XX DE Haemophilus high molecular weight protein Hmw1.
XX KW High molecular weight protein; Hmw1; protective vaccine; otitis;
XX KW sinusitis; bronchitis; Hib.
XX OS Haemophilus.
XX FN WO9421290-A1.
XX PD 29-SEP-1994.
XX PF 15-MAR-1994; 94WO-US002550.
XX PR 16-MAR-1993; 93US-00038682.

XX PA (BARE/) BARENKAMP S J.
XX PA (SGEM/) ST GEME J W.
XX Barenkamp SJ, St Geme JW;
XX WPI: 1994-316665/39.
XX N-PSDB; AAQ72293.
XX New immunogenic high mol. wt. proteins of non typeable Haemophilus -
XX useful in protective vaccines.
XX Claim 2; Page 31; 127pp; English.
XX The HmW1 protein encoded by this sequence is useful in a vaccine to
XX protect against disease caused by non-typeable Haemophilus which are not
XX controlled by H. influenzae type b (Hib) vaccines. The encoded protein
XX can also be used as a carrier for protective Hib polysaccharide (in a
XX conjugate vaccine against meningitis) or for other antigens, haptens, etc.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 1536 AA;
XX
XX Query Match 6.7%; Score 173.5; DB 2; Length 1536;
XX Best Local Similarity 22.8%; Pred. No. 0.0014;
XX Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
XX
XX 12 MKTHSTGLTVSNINAN-TIGYKQOQVFDLFSQDLAIGTSGQPNQAGNQAQVGS 70
XX 1011 IKAGVDGENSDATNNANLTIKTELK-----TQDLNI-----SGFNKAETAKDGS 1059
XX
XX 71 VRTIFTCGAFEPGNS-----VTDLAIGGKGFQVTLKDVHYTRAGNFRFTDQGLND 123
XX 1060 DLTIGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLSHKE--TSGSNNTIEDSSDNN 1116
XX
XX 124 PSQFTLMSRISNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
XX 1117 -AGLTIDAKNVTVNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAGTGS 1170
XX
XX 184 EANPYFALLEKWKNGT-----PISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSTG 238
XX 1171 ----ILGGIESSGSVTLTATEGALAVNIS-----GNTVTVTAN-SGALTTLA 1214
XX
XX 239 SKTEYLVAMPSPEDSGSASTDAGLL---MSTMTFSSNGELKNNATPTPGSATKOL 295
XX 1215 GSTINGTESVTTSSQSGDIGTISGGTVEVKATESLTQSNKIKATGTEANVTSATGTI 1274
XX
XX 296 NAWQAPLVNGLPQFSANFVCGAGIOTLDPGIKSQNMWAGAPASAAATGTDIGKLP 355
XX 1275 GGTISGNTVN-----VTNAGDLTVNGAE-----INATEGAATITSSGKL--- 1316
XX
XX 356 MPIOTSGNSTARNSSSTRYSODYPQGL---VDVITTSSEKL---QSK----- 401
XX 1317 ----TTEASSHITKAGQVNLSDAGSVAGSINAANVTLTGTLTVKGSNINATSGTL 1372
XX
XX 402 YSNQVDFVNIPIARFTSEGLRREGNNHVSATLDSGGPEFGIPGTSNYGKLSVNOLET 461
XX 1373 VINAKDAELNGAALGNHTVNAVNGSGSVIATSS---RVNITG---DLITINGL--- 1423
XX
XX 462 SNVDMSEVMNMIIQRFQNMKSVTTADTMLOKALELXR 502
XX 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQFGIASVDEIEAKR 1461
XX
XX RESULT 26
XX AAB01846
XX ID AAB01846 standard; protein; 1536 AA.
XX AC AAB01846;
XX XX
XX 12-SEP-2003 (revised)
XX 11-SEP-2000 (first entry)
XX

DE XX Haemophilus influenzae strain 12 HmW1A protein, SEQ ID NO:67.
XX XX HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
XX KW recombinant production; Escherichia coli; antibacterial; vaccine;
XX KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX KW detection; diagnosis.
XX OS Haemophilus influenzae; strain 12.
XX XX WO2000020609-A2.
XX PN 13-APR-2000.
XX PD 07-OCT-1999; 99WO-CA000938.
XX PF 07-OCT-1998; 98US-00167568.
XX PR 08-DEC-1998; 98US-00206942.
XX XX (CONN-) CONNAUGHT LAB LTD.
XX PA Loosmore SM, Yang Y, Klein MH;
XX PI WPI; 2000-303789/26.
XX XX N-PSDB; AAA52195.
XX DR Nucleic acid molecule for producing recombinant high molecular weight
XX PT proteins of Haemophilus which are used as a vaccine to provide protection
XX against Haemophilus induced diseases in humans.
XX XX Example 16; Fig 28A-Q; 307pp; English.
XX XX The invention relates to the recombinant production of Haemophilus
XX CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX CC expression construct used to effect recombinant expression comprises a
XX CC promoter functional in E. coli (e.g., the 17 promoter) operably linked to
XX CC a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
XX CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX CC and the hmwB and hmwC genes encode accessory proteins which are
XX CC responsible for post-translational processing and secretion of the HMW
XX CC proteins. The modified hmwABC operon used in the expression construct of
XX CC the invention contains an A gene modified such that it encodes only the
XX CC mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
XX CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX CC strains J95c, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
XX CC vectors are used for the production of recombinant H. influenzae HMW
XX CC proteins which can be used as vaccines to mediate a humoral or cell-
XX CC mediated immune response to provide protection against diseases in humans
XX CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and
XX CC tracheobronchitis). The HMW proteins are also useful as antigens in
XX CC immunoassays for detecting antibodies against Haemophilus, HMW proteins
XX CC and/or HMW peptides. The nucleotide sequences encoding the HMW proteins
XX CC can be used to isolate and clone hmw genes from other non-typeable
XX CC strains of Haemophilus via hybridisation reactions. The present sequence
XX CC represents an HMW protein from a non-typeable strain of H. influenzae.
XX CC (Updated on 12-SEP-2003 to standardise OS field)
XX SQ Sequence 1536 AA;

Query Match 6.7%; Score 173.5; DB 3; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.0014;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
XX
XX 12 MKTHSTGLTVSNINAN-TIGYKQOQVFDLFSQDLAIGTSGQPNQAGNQAQVGS 70
XX 1011 IKAGVDGENSDATNNANLTIKTELK-----TQDLNI-----SGFNKAETAKDGS 1059
XX
XX 71 VRTIFTCGAFEPGNS-----VTDLAIGGKGFQVTLKDVHYTRAGNFRFTDQGLND 123
XX 1060 DLTIGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLSHKE--TSGSNNTIEDSSDNN 1116
XX

QY 124 PSQFTLMGSRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSTDKTQS 183
DB 1117 -AGLTIDAKNVTVNNNITSEKAVSI-----SATSGEITTKTGTTINATGNVEITAQTS 1170
QY 184 EANPYFALLESGKNGT-----PPISTSNYSVAQPMRVYDQGNSHDITVYFDGAPSSGT 238
DB 1171 ----ILGGIESSGSVLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVANPSEDSAGSTDSAGLL---MSGTWFSSNGELKQWTAFTPTGSKTKDL 295
DB 1215 GSTIKGTESVTTSSQSDIGTISGTVVEKATESLTTOSNSXIKATTCGEANVTSATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQONMWAGAPASAAIGTIDIGKLPKM 355
DB 1275 GTTISGNTVN-----VTANAGDLTVNGAE-----INATEGAALTTSSGKL---- 1316
QY 356 MPIQTSSGNSTARNSSSTRYSQDGYPOGDL--VDVTITSEKL---QCK----- 401
DB 1317 ----TTEASHITSAKQVNLQAQDSVAGSINAANVTLTGTLTVKGSINATSGTL 1372
QY 402 YNSQVDFVFNIPARFTSEDLRRGNHYSATLDSGGPEFGLPGTSNYKLSVNOLET 461
DB 1373 VINAKDAELGAALGNHVTNATNANGSGSVIATTS---RVNITG-----DLITINGL-- 1423
QY 462 SNVDMSEVMNIIQRFQFQNSKSVTTADTMLQKALEKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQGIASVDEVIEAKR 1461

RESULT 27
ABU50020
ID ABU50020 standard; protein; 3705 AA.

AC ABU50020;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #35547.

EW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Yersinia pestis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PDDB; ACA53890.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 77944; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product; (7) identifying a compound that influences the activity of
CC or a gene on which the test compound that inhibits proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC on a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product's overexpressed; (11) a culture comprising strains in which the gene
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3705 AA;

Query Match 6.7%; Score 172.5; DB 6; Length 3705;
Best Local Similarity 21.7%; Pred. No. 0.0059;
Matches 119; Conservative 58; Mismatches 180; Indels 191; Gaps 24;

QY 3 GSLFICATGKTHSTGLTGVSNINANTIGYKQQQVVFQDLFSODLAIGTSGSGPNQA 62

DB 2359 GELIILGANGAFQTSLLNIASG--ASANINGVRO-----TVGAVNTGTVTTL 2403

QY 63 GMAQVGSVRIFTQGAPEGNSVTDLAIGKGFQVTLKDVHYTRAGNFRFTODGFLN 122

DB 2404 GNGELTGTDTLTNGTMINVTDGILNLENGGASSISGGL-----TNGILN 2449

QY 123 DPSG-FTLMGSRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSTDKT 181

DB 2450 IKGGDFTI-----SIDNN-----GLAQTNISDGASVT 2477

QY 182 QSEANPYFALLESGKNGTTPPISTSNYSVAQPMRVYDQGNSHDITVYFDGAPSSGT 241

DB 2478 -----LNGGTTIIGTN-----LGSSV 2494

QY 242 FEYLVANPSEDSAGS-----GT--DSAGLLMSGTWTFS-----SNGL- 283

DB 2495 IDVLGDLNLVADNSLANVISGDTTNTTATVILSGNSFSGAHQIQTNGELTVGQASNLG 2554

QY 284 AFTPTGSATKOLNAPQAPLVNGLPQFSANFVGAGIQPLTLD----- 325

DB 2555 A---SSATVNLGTLTSLHLNGVSESIAN-VLSGVAGSTVDIIGGADTALTANNSGFLG 2609

QY 326 -----FGIKSQONMWAGAPASAAIGTIDIGKLPMPQTSSGNSTARNSSSTR 376

DB 2610 QVALAGNSKLTVASTNNL--GASSVALAGT--GDTLSGNGTFGNSVTGSGVLQVT- 2664

QY 377 YSQDGYPOQDLVDVITITSEGLQKGYNSQVVDVFNIPL-----ARFT---SEDGLRRRG 428

DB 2665 -----DDAEVLTSSNGV---SNAVTTIDADATLNLDDIALFNHALTGGLNVA 2711

QY 429 NNHYSATLDSGGPEFG-LPGTSNYKLSVNOLETSNVDMSEVMNIIQRFQFQNSKSV 487

DB 2712 KNDASTAFDFGATVGGAFTGTVN-----LNNSTFDLSGNNNTT-VLAQATLKUSSGNL 2762

QY 488 TTADTMLQ 495

DB 2763 TSVGNGVQ 2770

RESULT 28
AAW30293
ID AAW30293 standard; protein; 1536 AA.
AC AAW30293;
DT 17-OCT-2003 (revised)
DT 14-APR-1998 (first entry)
XX
XX Non-typeable Haemophilus high mol. wt. surface protein HMW1.
XX
XX Non-typeable Haemophilus; high molecular weight surface protein; HMW1;
XX hmw1A gene; immunogen; vaccine; otitis media.
XX
XX Haemophilus influenzae; strain 12.
XX
XX Key Location/Qualifiers
FH Misc-difference 4 /note= "encoded by CTA"
FT Misc-difference 98 /note= "encoded by GAT"
FT Misc-difference 363 /note= "encoded by AAG"
FT
FT
XX W09736914-A1.
XX
XX
XX PD 09-OCT-1997.
XX
XX PF 01-APR-1997; 97WO-US004707.
XX
XX PR 01-APR-1996; 96US-00617697.
XX
XX PA (BARE/) BARENKAMP S J.
XX
XX PI Barenkamp SJ;
XX
XX WPI; 1997-503038/46.
DR N-PSDB; AAT90994, AAT90996.
XX
XX High molecular weight proteins of non-typeable Haemophilus influenzae -
XX useful for vaccine production.
XX
XX Claim 7; Page 66-70; 183pp; English.
XX
XX This protein comprises the high molecular weight surface protein HMW1
XX (125 kDa) of non-typeable Haemophilus influenzae strain 12 that has the
XX immunological ability to protect against disease caused by a non-typeable
XX Haemophilus strain and is characterised by at least one surface-exposed B
XX cell epitope that is recognised by monoclonal antibody AD6. The HMW1
XX amino acid sequence was deduced from the hmw1 gene sequence (see AAT90994
XX and AAT90996). The expressed protein is truncated, starting at residue
XX 442 of the full-length gene product. HMW2 (see AAW30294), HMW3 (see
XX AAW30291) and HMW4 (see AAW30292) have also been identified. A conjugate
XX comprising HMW1 linked to an antigen, haptan or polysaccharide, and a
XX synthetic peptide of 6-150 amino acids corresponding to at least
XX protective epitope of HMW1 are also claimed. HMW proteins, conjugates and
XX peptides can be used in vaccines, as immunogens for preparation of
XX antibodies and as antigens for detection of these antibodies. (Updated on
XX 17-OCT-2003 to standardise OS field)
XX
XX Sequence 1536 AA;

Query Match 6.5%; Score 167.5; DB 2; Length 1536;
Best Local Similarity 22.3%; Pred. No. 0.004;
Matches 116; Conservative 81; Mismatches 224; Indels 100; Gaps 23;
XX
XX 12 MKTSTGLTGVNNIANAN-TIGYKQVQVFDLFSODLAIGTSGGQNCAGQAGVGS 70
XX
XX 1011 IKAGVDGSDSDTANNANITIKTEKUL-----TQDLNI-----SGFNKAEITADGGS 1059
XX
XX 71 VRTFTQGAPEPGNS-----VTDLAIGKGFFQVTLKDVHYTRAGNFRFTQDGLND 123

Db 1060 DLITGNTSADGTNAKKVTFNQVDSKISADG-HKVTLSKVTETSGSNN---NTEHSSDN 1115
QY 124 PSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTOG 183
Db 1116 NAGLTIDAKNVTVNNITSHKAVSI-----SATSGEITTKTGTATTATNGVVEITAGTGS 1170
QY 184 EANPYFALLESMKNGCT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGT 238
Db 1171 ----ILGGIESSGSGVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVAWNPSEDGSAASGTDGALL---MSGTMTFSSNGELKNMTAFTGSAKDL 295
Db 1215 GSTIKTESVTSOSSGDIIGGTISGTVKATESLTTQSNISKIKATTGEANVTATGTI 1274
QY 296 NAMQPAPLVNGLPQFSANFVGAGIPLTLDGFIKQQNMWAGAPASAAAIGTDIGKLPMS 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPQTSNGSTARNSSSTRYSQDYPQCDL--VDVITSEGL---QGG----- 401
Db 1317 ----TTEASSHITSAGQVNLGSAQDSVAGSINAANVTINTTGTTLTVKGSNINATSGTL 1372
QY 402 YSNSQVDFYNTPLARFTSEDLRREGNNHYSATLDSGPEFGLPCTSNYKLSVNOLET 461
Db 1373 VINAKDAELGAALGNHTVVNATNANGSGSVIATSS---RVNITG---DLITINGL-- 1423
QY 462 SNVDSREMNMMIITQRFQPMNSKSVTTADTMQLKALEKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQGIASVDEVEIAKR 1461

RESULT 29
ABU38411
ID ABU38411 standard; protein; 2468 AA.
XX
XX AC ABU38411;
XX
XX DT 19-JUN-2003 (first entry)
XX
XX DE Protein encoded by Prokaryotic essential gene #23938.
XX
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX OS Pseudomonas aeruginosa.
XX
XX PN W0200277183-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 21-MAR-2002; 2002WO-US009107.
XX
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX
XX DR WPI; 2003-029926/02.
XX
XX DR N-PSDB; ACA42281.
XX
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX PS Claim 25; SEQ ID NO 66335; 1766pp; English.
XX
XX CC The invention relates to an isolated nucleic acid comprising any one of

the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:	
(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> , <i>X. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
CC	ftp.wipo.int/pub/published_pct_sequences
CC	Sequence 2468 AA;
SQ	
Query Match 6.4%; Score 166; DB 6; Length 2468;	
Best Local Similarity 21.9%; Pred. No. 0.01;	
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;	
QY	20 GTVNNIANANTIGYKQQQVWFQDLFSQDLAIGSTGSG-----PNOAGMGAQVGSV 71
DB	1298 GTVNAVA-----QDPA-GNTGPGQSTTVDAVAPNTPVNPNSGNL 1337
QY	72 RTITQGAPEPNSVTDLAIGGKGFQVTLKDVHYTRAGNFRFTQDGFNDPDSGFTLMG 131
DB	1338 ----LNGTAEPGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSQLPNGTVNVNVA 1387
QY	132 SRISSNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKDQSEANPYFAL 191
DB	1388 SDAAGNTSLPAT-----TVDSLSLPIQVDPNSGVISGTADAGNTIIITDNGNPIGQV 1443
QY	192 LESWKNG--TPPISTSNYSYAQPMRYVDQGNSHDITVYFDG-APSGTSGKTFEYLAVM 248
DB	1444 TADGSGNWSFTPGIPLDGTVVNVVARSNVDSPAPAVITVDGVAPAAP-----VI 1494
QY	249 NPSEGGNASTGDSAGLLM-----SGTWTF-----SNGELKNMTAFT 286
DB	1495 DPS-NGTEISGTAEAGATVILTDGGNPIGQATADGSGNWTFTPTPLANGTVINVAQD 1553
QY	287 PTG-----SATKDLNAWQAPLVN-----GLPQFSANFV-----GAGIPLTLDIFG 327
DB	1554 PAGNTSGPASVTVDAIA-PPAPVNPNSGVISGTAEAGATVILTDGNGNPIGQVTAAD-- 1610
QY	328 IKSQONWAGAPASAAIGTDIGKL-----PSMFIQTSSGNSTA 367
DB	1611 ---GSGNWSFTPGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAAPVIDPNSGVIA 1667
QY	368 RNGSSSTRYSQD--GYPOGLVDVTITSEKLGQKYSNQSVDFYNIPLARFTSEDLGR 425
DB	1668 GTAEGATVILTDGNGNPIGQ-----VTADG--SGNWSFTPGTPLSNGTVNVAQDA-- 1718
QY	426 REGNNH--YSATLDSGGGPEFLPGTSNYKGLS 455
DB	1719 -AGNTSGVSTTVDAVAPATPVIDPNSGVELS 1749

XX	RESULT 30
XX	ABP59933
ID	ABP59933 standard; protein; 2468 AA.
XX	XX
AC	ABP59933;
XX	XX
DT	28-AUG-2003 (first entry)
DE	Microbial resistance gene PA1874 protein.
DE	XX
KW	Biofilm; microbial resistance.
XX	Unidentified.
OS	WO2003041483-A2.
PN	22-MAY-2003.
XX	XX
PD	18-SEP-2002; 2002WO-US029565.
XX	XX
PF	18-SEP-2001; 2001US-0323241P.
XX	XX
PR	(DART-) DARTMOUTH COLLEGE.
XX	XX
PI	O'toole GA, Mah T;
XX	XX
XX	WPI; 2003-468567/44.
DR	N-PSDB; ACC59398.
XX	XX
PT	Identifying modulators of microbial resistance of organisms in biofilms, e.g. inhibitor of biofilm formation, by employing expression controls, or efflux pumps containing polypeptides, of genes associated with biofilm resistance.
PT	XX
XX	Claim 33; Fig 7; 102pp; English.
PS	XX
CC	The present invention relates to a method of identifying a compound capable of altering the sensitivity of a microorganism to an antimicrobial agent by employing efflux pumps comprising polypeptides encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143, PA2390 or PA1163. The method is useful for identifying modulators of microbial resistance of an organism in a biofilm. The methods are also useful for identifying genes that encode proteins that play a role in biofilm resistance. The method is particularly useful for screening compounds or discovering compositions that will inhibit biofilm formation and overcome their resistance mechanisms. These methods are particularly useful in medical, industrial or natural settings, where formation of biofilms can have serious negative consequences and result in high costs both in human health and economic terms. The present sequence is a protein shown in the exemplification of the invention
CC	Sequence 2468 AA;
SQ	
Query Match 6.4%; Score 166; DB 6; Length 2468;	
Best Local Similarity 21.9%; Pred. No. 0.01;	
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;	
QY	20 GTVNNIANANTIGYKQQQVWFQDLFSQDLAIGSTGSG-----PNOAGMGAQVGSV 71
DB	1298 GTVNAVA-----QDPA-GNTGPGQSTTVDAVAPNTPVNPNSGNL 1337
QY	72 RTITQGAPEPNSVTDLAIGGKGFQVTLKDVHYTRAGNFRFTQDGFNDPDSGFTLMG 131
DB	1338 ----LNGTAEPGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSQLPNGTVNVNVA 1387
QY	132 SRISSNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKDQSEANPYFAL 191
DB	1388 SDAAGNTSLPAT-----TVDSLSLPIQVDPNSGVISGTADAGNTIIITDNGNPIGQV 1443
QY	192 LESWKNG--TPPISTSNYSYAQPMRYVDQGNSHDITVYFDG-APSGTSGKTFEYLAVM 248
DB	1444 TADGSGNWSFTPGIPLDGTVVNVVARSNVDSPAPAVITVDGVAPAAP-----VI 1494

QY 249 NPSGDSRAAGTSDSAGLMM-----SCTMTFS-----SNGELKNMTAFT 286
 DB 1495 DFS-NGTEISGTAAGATVILTDGGNGPIGQATDGGNGWTFSTPLANGTVINAVAQD 1553
 QY 287 PTG-----SATKOLNAPQAPLVN-----GLPOFSANFV-----GAGIOPLTLDFG 327
 DB 1554 PAGNTSGPASVTVDAIA-PPAPVNPNSGWSGTAEAGATVILTDGNGNPIGQVTD-- 1610
 QY 328 IKSQNMWAGAPASAAATGTDICKL-----PMMPIQTSGNSTA 367
 DB 1611 ---GSGNWSFTGTPLANGSVINALAQDAAGNNSFTSATVDSLAPAAPVIDPSNGSVIA 1667
 QY 368 RNGSSSTRYSQD--GYQGDLDVDTITSEGKLOGKYSNSQVDFYNIPLARFTSEDLGR 425
 DB 1668 GTAAGATVILTDGNGNPIGQ-----VTADG--SGNWSFTGTPLSNGTVVNAVAQDA-- 1718
 QY 426 REGNHH--YSATLDSGGPEFGLPSTNYGKLS 455
 DB 1719 -AGNTSGPVSTTVDAVAPATPVIDPSNGVELS 1749

RESULT 31
 AAW20361
 ID AAW20361 standard; protein; 124 AA.
 AC AAW20361;
 DT
 XX
 XX
 XX
 DE H. pylori flagella-associated protein 26508588.aa.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 27 /label= unknown
 FT /note= "encoded by TWA"
 FT Misc-difference 58 /label= unknown
 FT /label= unknown
 FT /note= "encoded by AGS"
 XX
 PN WO9640893-A1.
 XX
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US009122.
 XX
 XX
 PR 07-JUN-1995; 95US-00487032.
 PR 01-APR-1996; 96US-00630405.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Smith D, Berglindh OT, Wellgaard BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67553.
 XX
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 XX
 PS Claim 56; Page 550; 1481pp; English.
 XX
 CC This sequence is a H. pylori flagella-associated protein. The protein may
 CC be used in a vaccine to prevent or treat H. pylori infection or to
 CC identify H. pylori polypeptide binding compounds, useful as potential H.
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by

CC mechanically shearing the bacterial DNA. The sequences were analysed for
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF were
 CC analysed for significant homology to other known or exported membrane
 CC proteins. Having identified and determined the sequences of interest,
 CC particular regions can be isolated from H. pylori by PCR amplification
 CC for recombinant polypeptide production, e.g. in E. coli hosts
 XX
 XX Sequence 124 AA;
 QY
 Query Match 6.4%; Score 165.5; DB 2; Length 124;
 Best Local Similarity 32.0%; Pred. No. 0.00016;
 Matches 39; Conservative 24; Mismatches 48; Indels 11; Gaps 3;
 QY 392 ITSEGKLOGKYSNSQVDFYNIPLARFTSEDLRREGNNHY---SATLD-----SGGPE 442
 DB 3 IEENGVISLAFNGVVEPVARIGIXAFTNDQGLRKLGNLYEMOEGTINGENRPLXGNPI 62
 QY 443 FGL--PGTSNYGKLSVNOLETNSVDMSEVMNMIIQRGFQNSKSVTTADTLMQALEL 500
 DB 63 LGWDEEGKLGKIRKHYLETNSVNAGNALTNLILMQRYSMNARAFAGGDDMIKEAISL 122
 QY 501 KR 502
 DB 123 KK 124

RESULT 32
 ADC01014
 ID ADC01014 standard; protein; 5291 AA.
 AC ADC01014;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 1058.
 XX
 KW enterohaemorrhagic; anti-bacterial.
 XX
 OS Escherichia coli; 0157:H7.
 XX
 PN JP2002355074-A.
 XX
 PD 10-DEC-2002.
 XX
 PF 24-JAN-2002; 2002JP-00015959.
 XX
 PR 24-JAN-2001; 2001JP-00112010.
 XX
 PA (UYTS-) UNIV TSUKUBA.
 XX
 DR WPI; 2003-451640/43.
 XX
 FT Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule
 FT and a polypeptide and its use, a polypeptide, a vector and a host cell.
 XX
 PS Claim 3; SEQ ID NO 1058; 2067pp; Japanese.
 XX
 XX The invention relates to a novel enterohaemorrhagic Escherichia coli
 CC 0157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 CC has anti-bacterial activity. The polypeptide can be used in detection
 CC and/or treatment of 0157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic E. coli 0157:H7 was determined. The present
 CC sequence represents an E. coli 0157:H7-specific polypeptide of the
 CC invention.
 XX
 XX Sequence 5291 AA;
 QY
 Query Match 6.4%; Score 165; DB 7; Length 5291;
 Best Local Similarity 22.6%; Pred. No. 0.036;
 Matches 145; Conservative 71; Mismatches 189; Indels 236; Gaps 35;

PS	Claim 8; Page 301-306; 2054pp; English.
XX	This sequence represents a Helicobacter pylori GHPO protein of the
CC	invention. The polypeptides can be used for preventing or treating
CC	Helicobacter infections, and gastroduodenal diseases associated with
CC	these infections, including acute, chronic, and atrophic gastritis, and
CC	peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC	used for the production of antibodies. The products can also be used for
CC	detection and diagnosis
XX	
SQ	Sequence 1230 AA;
	Query Match 6.2%; Score 160; DB 2; Length 1230;
	Best Local Similarity 21.6%; Pred. No. 0.011;
	Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;
Qy	20 GTVSNNTANANTIGYKQQQV--FDLFSQDLAIGS-TGSQGPNQAGQAQVGSVRIET 76
Db	542 GTTINTQAKSNAPKLKAMVNVNNEEAKTANLAQSSGTTTQSPNSTVMGA-----LNTVL- 596
Qy	77 QGAPEPGNSVTDLAIGKGFQVLEDKVHYTRAGNERFT--QDGLNDPQSGFTLMGSRIS 135
Db	597 -----QNVSNFQCSQIQAFOQESNIQAWANAY 625
Qy	136 NNPNIKKETLEP-----IQLDENDPTVAKS-PAKTSTALNAVYNLGDSTDKTQSEANP 187
Db	626 NTNGSQSQEPTNNNQDLRIQLRANFYQLINTINQOVPTDMNALINQSQOQTQSGSASN 685
Qy	188 YFALLSEWKNGTPIISTNSVSYAQPMRVYDQGNSHDITVYFDGAPSTGSKTFEYIVA 247
Db	686 NNACASMSG-----SNGNWCYQQ-----WSDSKAYYSGLQSALGYQT----- 723
Qy	248 MNPSEDCSAAGSTDAGILMSGT--MTFSSNGELK-----NMTAFTPTGSGATK 293
Db	724 -----QANTQSGSNGCHSITVWQIILTSNGLLNQIITNLKSVNNGNGASGTSGNGTS 778
Qy	294 DLNAWQAPLVNGLPQPSANFVGAGIQPLTLDFGIKSQQNMWAGAPASAAAIGTDIGKLP 353
Db	779 QIN-----TAYQMLT-----DASDQKLG----- 797
Qy	354 SMPPIQTSSG-----NSTARNSSSTR-RYSQDGYPOGDLVDVITITSEKLGQKYSN 404
Db	798 -----YSSSSGSGNGYTPCNSI--NGSNKTSNGCYEPNKQNTATATDTSNLQKVND 851
Qy	405 SQVYDFVNIPIARFTS---EDGLRR--EGNNHYSATLDS--GGPEFGLPGTSNYGKL--S 455
Db	852 AQKI--ANILASSGNKGVENGALQFPEALKKNSSLSNLKCGSSGSGSGITCGWLINL 909
Qy	456 VNQLETGNDVMSREWNMI-----IQRGP-QVNSKSVTTADTLMQKAL 498
Db	910 LGAIPTNGVSDTNLNLILLTFIKTAGFIQNDDSVSTSLSAFOAI 956
RESULT 36	
ABU30969	
ID	ABU30969 standard; protein; 1230 AA.
XX	
AC	ABU30969;
DT	19-JUN-2003 (first entry)
XX	
DE	Protein encoded by Prokaryotic essential gene #16496.
XX	
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design..
OS	Helicobacter pylori.
XX	
FN	WO200277183-A2.
XX	
PD	03-OCT-2002.
XX	
PF	21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362999P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX WPI; 2003-029926/02.
XX DR N-PSDB; ACA34839.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PT
XX
XX Claim 25; SEQ ID NO 58993; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC <http://wipo.int/pub/published/pct/sequences>

XX	SQ	Sequence	1230 AA;
		Query Match	6.2%; Score 160; DB 6; Length 1230;
		Best Local Similarity	21.6%; Pred. No. 0.011;
		Matches	114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;
Qy	20	GTVSNINANTIGYKQQVV--FODLFSODLAIGS-TGSGPNQAGMGACVGSVRTFT	76
Dd	542	GTITNTQAKNAPKLKAMVVNNHEEAKTANLAQSSGTTTQSPNSTVMGA---LNIVL-	596
Qy	77	QGAFEPGNSVTDLAIGCKGFQVLTEDKVHYTRAGNFRFT-QDGFLLNDPSGFTLMGRIS	135
Dd	597	-----QNVSNEPQQSIQNAFOQESNIQAWAANAIV	625
Qy	136	NNPNIKKETLEP-----IQLDENDPTVAKS-PAKTSTALNAVNLGDSTDTKTQEASNP	187
Dd	626	NNTGSGSQEWTPNNQDLRLQLRANFFQLINTINQOVPDMNALINQSQQQTCTSGSASN	685
Qy	188	YFALLSEWKNGTPTPISTSYNSYAOPMRVVDQGNSHDITYVFQCAPSSTCSKTFEYLVA	247
Dd	686	NNACASGMGG-----SNGWNCYQQ-----WSDSKAYISGLQSALEYQT-----	723

	6.2t	Score 160;	DB 6;	Length 1230;
Query Match				
Best Local Similarity	21.6t	Pred. No. 0.011;		
Best Global Similarity	21.6t	Pred. No. 0.011;		
Number of Matches	71	Matches 182.	Indels 160;	Gaps 25

QY	20	GTVSNNIANANTIGYKQQVV--FQDLFSODLAIGS-TGSGPNQAQGAQVGSVRTFT	76
Db	542	GTITNTQAKSNAPKLKAVVVNNEEEAKTANLAQSSGTTTQSPNSTVWGA---LNTVL-	596
QY	77	QGAPEPGNSVTDLAIIGKGFQVLTDEKVHYTRAGNFRFT-QDGFINDPSGFTLMGSRIS	135
Db	597	-----QNVSNFQQSIGNAFQESNIQIAWANAIV	625
QY	136	NNPNIKKETLEP-----IQDLFNDPTVAKS-PAKTSTALNAVNLGDSDTKQSEANP	187
Db	626	NTNGSQSQEMTPNNQDLRLQLRANFYQLINTINQVPTDMNALINQSQOQTQSGSASN	685
QY	188	YFALLGSWKNGTPTPITSNYSYAQPMRWVDQGNSHDITVYFDGAPSSGTSKTFEYLVA	247
Db	686	NNACASGMG-----SNGWICYQ-----WSDSKAYISGLGSAALGYQT-----	723

XX
DE protein encoded by prokaryotic essential gene #16496.

XX

XX
XX

XX
XXXXXXXXXXXXX
XXXXXXXXXXXXX

FN
XX
WJZ00277183-AZ.

PD 03-OCT-2002.
yy

PF 21-MAR-2002; 2002WO-US009107.

ID	AAW73072 standard; protein; 270 AA.
XX	04-MAY-2000 (first entry)
XX	C. jejuni flgF protein.
XX	Basal body rod protein; flgF; flagellum; vaccine; immunogen;
XX	pathogenic bacteria; detection; antibacterial.
XX	Campylobacter jejuni.
XX	US6020125-A.
XX	01-FEB-2000.
XX	07-JUN-1995; 95US-00483857.
XX	08-MAY-1995; 95US-00436748.
XX	(CONN-) CONNAUGHT LAB LTD.
XX	Louie H, Chan VL;
XX	WPI; 2000-146875/13.
XX	N-PSDB; AAZ88553.
XX	Recombinant basal body rod protein producible by a transformed host for use in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens.
XX	Claim 1a; Col 21-24; 28pp; English.
XX	This invention describes a novel recombinant basal body rod protein (I) producible by a transformed host containing an expression vector comprising a nucleic acid selected from: (a) an entire nucleic acid sequence of 1800 base pairs (bp), or the nucleic acid sequence of the flgF gene having 810 bp, both given in the specification; (b) a nucleic acid encoding the amino acid sequence of the flgF protein, a 270 residue sequence, given in the specification; (c) a nucleic acid sequence encoding a functional flgF basal body rod protein of a flagellum of a strain of Campylobacter; or (d) an immunogenic fragment of an flgF protein of (a), (b) or (c), and expression means operatively coupled to the nucleic acid molecule for expression by the host of a basal body rod protein of a flagellum of a strain of Campylobacter. (I) is useful in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens that produce basal body rod proteins. (I) are also useful as immunogens, as antigens in immunoassays, or for procedures for the detection of antibodies. The Campylobacter, basal body rod protein and/or peptide antibodies. The product of the invention has immunostimulatory activity. This sequence represents the Campylobacter jejuni flgF protein described in the method of the invention
XX	Query Match 6.0%; Score 155; DB 2; Length 270;
XX	Best Local Similarity 30.7%; Pred. No. 0.003;
XX	Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;
XX	1 MNGSLFICATGKMTHTSLGTGVSNNANANTIGYKQOQV---FQDLFSQ---DLAIGS- 53
XX	1 MONGYYQATGGMVTFQNKLDVTNNLANINTSGYKRDVDVIADFKRIFKETQDELPIENH 60
XX	54 --TGSGPQAGMG-AQVGSVRTIFTQGAPEPNSVTDLAIGSGKFQV--TLEDKVHYT 108
XX	61 TRDASRFVNTTIDGIPQVQSQEYTFSLGSLKATNPLDLAMTRDAPFLVQTKDGEVRLT 120
XX	109 RAGNFRFTQDGFNDPSPGFTLMGSRISNNP 138
XX	121 KGNFQDLDEGLVKNQGYKVLSSDYFNNP 150
XX	RESULT 39
XX	AAW51376
XX	ID AAW51376 standard; protein; 270 AA.
XX	AAW51376;

ID ABU36440 standard; protein; 2204 AA.
AC ABU36440;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #21967.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Mycobacterium tuberculosis.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI WPI: 2003-029926/02.
DR N-PSDB; AC40310.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 64364; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2204 AA;

Query Match

5.9%; Score 151.5; DB 6; Length 2204;

Best Local Similarity 20.2%; Pred. No. 0.11;
Matches 119; Conservative 64; Mismatches 220; Indels 185; Gaps 24;
QY 3 GSLFIGATGKTHSTGLGTVSNANANTIGYKQOVWFQDLFSODLAIGSTG----- 55
DB 1458 GNIGIGITG--TQIGFGSFSNHNIGLFGSDGNGVGFNSGTGNVGLNTGTANFGIA 1515
QY 56 SQPNAGAGCAQVGSVRTTFTQGAPEPPGNSVTDLAIGGK-----GFFQV----- 99
DB 1516 NSGSFNTGLG-NTGSTNT-----GLFNPNGVNTGVGNTGTSINTGTSFNTGSTNTGSF 1570
QY 100 TLED-----KVHYTRAGNRF-----TQDGF----- 120
DB 1571 NLGDHNTGSFNSGDYNTGFNAGDYNTGVANTGNTGAFISGNYSNGFFFWGDDYQGLIG 1630
QY 121 -----LNDPSGFTLMGSRISNNPN-----IKKETLEP 147
DB 1631 LSTTITIPBIPRYDLSVPIDIPITGTVVATTNSFTTIPGFOIRVLLGPAAVLVNMIGP 1690
QY 148 IOLDENDPTVAKSPAKTSTALNAVNLGSDTDKTOSEANPYFALLESKNGTPTPISN 207
DB 1691 ITIDNVQVIAIDSPICQTISMVGTGGFGPIPIGIGISGTPGF-----GNST----- 1736
QY 208 YSAQPMRYDQGNSHDITVYFDGAPSTSTGSKTPEYLAMNPSDGSAAAG-TDSAGLL 266
DB 1737 ---TGPSSGFFHTGAGVSGFGNFCAGNMSGNF-----GAGNSGFFNAGGLG 1782
QY 267 MSGTWTFS-NGELKNMTAFTPTGSATKDLNAPLIVNGLPQFSANFVGAGIQ---PL 322
DB 1783 NSGLNFGALQSLANLGNNTISGVYNTTLDLTPA-FGSGIANIGANLAGLFDNTGNL 1841
QY 323 TLDFGIKSOQNMWAGAPASAAAIGTDIGKLPWMPIQTSSGNSTARNSSSTRY----- 377
DB 1842 TLNFGVANQGLNAG-----IGNLGSVNI GFVNTGDSNLGIGNLGLNFGGWN 1890
QY 378 ---SQDGYQGLDLDVTTTSEKLOKQYKN--SQVDF-----YNIPLARFTSD----- 422
DB 1891 GGNNGIANTGTFIDGLANLGSYNIGLANLGDNDLGNAGSYNIGFANFGDNLGFANT 1950
QY 423 -----GLRREGNNHYS-----ATLDSGGPEFGL--PGTSNYG 452
DB 1951 GSYNIGFANTGNNIGVGLTGNQGIGIGISLNSGNNIGLFSNGSGNIG 1998

Search completed: October 26, 2004, 09:05:38
Job time : 81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2004, 20:00:04 ; Search time 484 Seconds

(without alignments)
5318.298 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGLFVIGATGKTHSTGLG.....NSKSVTTADTMLQALELKR 502

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters: 6826950

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.spool/US10009823/runat_26102004_100131_1718/app_query.fasta_1.647
-DB=Published Applications NA -QFMT=fastp -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCUT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10009823 @CGM_1_1_354 @runat_26102004_100131_1718
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgm2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgm2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgm2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgm2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgm2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgm2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgm2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgm2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgm2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgm2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgm2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgm2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgm2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgm2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgm2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgm2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgm2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 19: /cgm2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgm2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgm2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2586	100.0	1509	15	US-10-009-823A-2
2	529.5	20.5	1242	16	US-10-282-122A-11957
3	527.5	20.4	1389	15	US-10-127-032-16
4	506.5	19.6	1239	16	US-10-282-122A-14579
5	485	18.8	2157	16	US-10-335-977-58
6	484.5	18.7	1163020	16	US-10-398-221-10
7	484.5	18.7	3011208	16	US-10-398-221-2058
8	469.5	18.2	5998	16	US-10-398-221-3855
9	414.5	16.0	640681	9	US-09-790-988-1
10	296	11.4	665	16	US-10-335-977-56
11	288	11.1	789	16	US-10-335-977-1
12	288	11.1	816	16	US-10-335-977-2
13	288	11.1	837	16	US-10-335-977-3
14	237	9.2	555	9	US-09-974-300-1888
15	215.5	8.3	1524	9	US-09-974-300-1846
16	201.5	7.8	459	16	US-10-335-977-78
17	193	7.5	5445	14	US-10-210-296-2
18	193	7.5	5445	15	US-10-449-462-2
19	189.5	7.3	1395	16	US-10-335-977-82
20	180.5	7.0	1965	15	US-10-369-493-28400
21	180.5	7.0	1965	15	US-10-369-493-31160
22	173.5	6.7	3285	15	US-10-193-764-64
23	173.5	6.7	5116	13	US-10-092-880-1
24	173.5	6.7	5116	15	US-10-193-764-62
25	173.5	6.7	9171	13	US-10-092-880-5
26	172.5	6.7	11118	16	US-10-282-122A-41760
27	166	6.4	273	16	US-10-335-977-55
28	166	6.4	7407	15	US-10-246-330-3
29	166	6.4	7407	16	US-10-282-122A-30151
30	165.5	6.4	372	16	US-10-335-977-77
31	165	6.4	25165	14	US-10-114-170-39
32	164	6.3	1029	16	US-10-335-977-57
33	162	6.3	5863	16	US-10-147-299A-3
34	160	6.2	3693	16	US-10-282-122A-149
35	160	6.2	3793	9	US-09-881-752A-149
36	156.5	5.9	9219	16	US-10-282-122A-12963
37	153.5	5.9	2037	15	US-10-369-493-37512
38	152.5	5.9	554	16	US-10-398-221-1330
39	151.5	5.9	810	16	US-10-335-977-17
40	151.5	5.9	6615	16	US-10-282-122A-28180
41	150	5.8	3396	15	US-10-193-764-60
42	149.5	5.8	3543	15	US-10-193-764-58
43	149.5	5.8	3568	15	US-10-193-764-58
44	148.5	5.7	819	16	US-10-335-977-18
45	147.5	5.7	558	16	US-10-335-977-79

ALIGNMENTS

RESULT 1

US-10-009-823A-2
; Sequence 2, Application US/10009823A
; Publication No. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Ankenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES
; FILE REFERENCE: DAVI150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Lawsonia intracellularis

US-10-009-823A-2

Alignment Scores:

Pred. No.: 1,05e-259 Length: 1509
 Score: 2586.00 Matches: 502
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-10-009-823A-1 (1-502) x US-10-009-823A-2 (1-1509)

QY 1 MetMetGlySerLeuPheThrGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 1 ATGATGGGAGTTTGTATTGTCACAGATAGAAAACCCATAGTACAGGGTGGGT 60
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 61 ACTGTCTCCATAATATTGCTAACGCAATACCATTTGGTATATAGCAGCACAGTAGT 120
 QY 41 PheGlnAspLeuPheSerClnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 121 TTTCAGACCTGTTTGTAGTCAAGATTAGCAATAGGTTCTACTGGAAGTCAAGGGCCAAAC 180
 QY 61 GlnAlaGlyMetGlyValaGlnValGlySerValArgThrIlePheThrGlnGlyValaPhe 80
 DB 181 CAGGCTGTATGGGAGCACAGGTTGGAGTGTTCGCACATTTTACACAGGGTGGCTTT 240
 QY 81 GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPheGlnValThr 100
 DB 241 GAACCTGGCAATAGTGTAAACAGATCTTGTATTTGGTGGAAAGGTTTTTTTTCAGGTTTACA 300
 QY 101 LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe 120
 DB 301 TTAGAGTAAGTACACTATACAGCAGCAGGAGGAAATTTTGTCTTACTCAAGATGGTTTT 360
 QY 121 LeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIle 140
 DB 361 TTAATGATCTCAGCGATTTTACTTTTAAATGGCTCAAGATATCTAATAATCCTAACATA 420
 QY 141 LysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer 160
 DB 421 AAAAAGGAAACCTTGAAACCAATTCAGTTAGACTTTAATGATCTCAGTAGCAAGTCT 480
 QY 161 ProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLys 180
 DB 481 CCTGCAAAACACAGTACAGCATTAAACGCTGTGGTAAACCTTGTGTAGTAGTACAGATAAA 540
 QY 181 ThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr 200
 DB 541 ACACAAAGTGAAGCTAATCCATATCTTTCACCTTCTTGAGAGCTGGAAAGAAATGGAACA 600
 QY 201 ProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln 220
 DB 601 CCTCCTATTCTACATCAAACTACTCATATGCACAACTATGAGATATATGATCAACAA 660
 QY 221 GlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLys 240
 DB 661 GGAAATTTCTCAGGATATPACTGTATATTTTGTATGGAGCACCTCTTCAACAGGAAGTAAA 720
 QY 241 ThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaLysSerGlyThr 260
 DB 721 ACATTGAAATACCTTGTAGCTATGATCTTGTAGTGAAGTGAAGTGTCTGATCAGGAACA 780
 QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 781 GATAGTGCAGGCTCTCTTAATGTCTGGAACCTATGACATTTTCAAGTAATGCGCAATTTAAA 840

QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
 DB 841 AATATGACAGCTTTTACTCTCTACTGGCTCTGCACAAAAGATTAAATGATGGCAACA 900
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln 320
 DB 901 GCACCATTTAGTCAATGGTTTACCACAGTTTTCAGCAAAATTTTGTGGTGCAGGAATACAG 960
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 DB 961 CCTTTAACAATTAGACTTTTGAATTTAAAGCCAAAGATATGTGGCAGAGCTCCAGCA 1020
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 1021 TCCGCTGCTGCCATAGGTACAGATATTGGAAATTTGCCATCAATGATGCCAATACAAACA 1080
 QY 361 SerSerClyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 DB 1081 TCCAGCGGTAATTTCTACAGCAAGAAATGGATCATCTTCAACAAGAGATATATAGCCAGAT 1140
 QY 381 GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly 400
 DB 1141 GGTATCTCTCAGGAGATCTAGTAGTGTCAATTTACCTCTGAAGGGAATTTACAGGT 1200
 QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
 DB 1201 AAGTATAGTAATAGTACAGTCTGTTGATTTTATAATATTTCTTTAGCACGCTTTACAAGT 1260
 QY 421 GluAspGlyLeuArgArgGluGlyValAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
 DB 1261 GAGGATGGATTAAAGCAGAGAGGGAATTAACCATTTATTCGCAACACTTGACTCAGTGG 1320
 QY 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
 DB 1321 CCAGAGTTTGGATTGCCAGGAACATCTAACTATGAAAACTTAGTGTGAATCAACTTGAG 1380
 QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
 DB 1381 ACTTCTACGTAGCATGAGCAGAGAAATGGTTAATATGATTATTATTCAACGTGGTTTT 1440
 QY 481 GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
 DB 1441 CAGATGAATAGTAATCTGTGTACACAGCAGACACAATGCTACAAAAGCACTTGAACATA 1500
 QY 501 LysArg 502
 DB 1501 AAGCGT 1506

RESULT 2

US-10-282-122A-11957

; Sequence 11957, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR FILING DATE: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 11957
;; LENGTH: 1242
;; TYPE: DNA
;; ORGANISM: Burkholderia cepacia
US-10-282-122A-11957

Alignment Scores:
Pred. No.: 3,14e-45 Length: 1242
Score: 529.50 Matches: 149
Percent Similarity: 44.20% Conservative: 72
Best Local Similarity: 29.80% Mismatches: 182
Query Match: 20.48% Indels: 97
DB: 13 Gaps: 13

US-10-009-823A-1 (1-502) x US-10-282-122A-11957 (1-1242)

Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAniLeuAla 27
Db 16 GGCTTGAGCGGATGGCGGTGCGTGCAGCAATCTCGACGTATCGGCACAAATCGCG 75
Qy 28 AsnAlaAsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGln 47
Db 76 AACGCAACACGGTCGCTTCAACGACGAGCGCGCAACTTCGCGGACATGTACGCGAAT 135
Qy 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
Db 136 TCGGTCCGCGAGTCGCTCAAC-----ACGCAGATTCGCGCATCGGACGCGG 180
Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
Db 181 CTCGCTCGGTGCGAGCAATTCGCGCAGGAGCGATCAATTCGACGAAGTCGTGCTC 240
Qy 88 AspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr 107
Db 241 GACGTGCGATCAACGCGCAACGGCTTCTTCAGATGTCAGCAACGCGCGTGACCACTAC 300
Qy 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
Db 301 TCGCGGACGCGACGTTCCATCGGCACAGAACGCGCGCATGCTCGACGCGCAGGCGCG 360
Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro 147
Db 361 AACCTGATGGCTATGCGCGCGCGCGCGCGCGGTGATCAACACCGCGCAGACCGTGGCG 420
Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
Db 421 CTCACGCG-----CCGACCAACAACTTCGCGCGCGCGCGCGCGCGCGCGCGCG 465
Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
Db 466 ATCCCGCGCGAGTCAACTCG-----AACGCGCAGCAGCAAGGTGCGCGCGCAAG----- 513
Qy 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsn 207

514 -----ACGCCGTTCAACGGCAGCGACCAAC 537
Qy 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225
Db 538 AGCAGGTACAACTCACTCGTCGATCCAGGTCTACGACAGCTCGCGGGTTCGACGAG 597
Qy 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
Db 598 GTCACGATGTACTTC-----GCGAAGAGCGCGCGCGCGCAG----- 633
Qy 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
Db 634 -----TGGCTGGCCCTACGCGCGCGGTGCGAGCGCGCAGCGCGCG 669
Qy 266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 670 ACGAATCTCGGCACCGTTCAGTTCGTCGCGCGCGGATCAGCTCGACGACGTCG----- 726
Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
Db 727 GCGCGCAGCGCG-----CAGCCGACGCGCG----- 750
Qy 306 GlyLeuProGlnPheSerAlaAsnPheVal-----GlyAlaGlyIleGlnPro 321
Db 751 AGCTTCGCGCAGTTCGCGTTCGATCCGGAACAGCGCGCGCGCAATCCGCGAAG 810
Qy 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSer 341
Db 811 CTGACGCTCGACCTG----- 825
Qy 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
Db 825 ----- 825
Qy 362 SerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgTyrSerGlnAspGly 381
Db 826 -----GGCGGACGACGACGACGCGCGCAAGGACGCGGTGACCAATCTCGCGCAGCGCGC 882
Qy 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 883 TTCGCGAGCGGACGCTGACGAGCTTCGATCGCGCAGCGCGCGCAAGCTGACCGCGAAC 942
Qy 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 943 TACTCGACGCGCAGACGCGCGGTGCTCGCGCTGATCGCGCTCGCAACTTCAACACCGC 1002
Qy 422 AspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyPro 441
Db 1003 AACCGGCTCGTGAACATCGCGCGCAACCAAGTATGCGGAACGCGCGCGTCCGCGCGCG 1062
Qy 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 1063 CAGATCG 1122
Qy 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGln 481
Db 1123 TCGAAGCTGAACCTGACCG 1182
Qy 482 MetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1183 GCGAAGCGCGCAGCATCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242

RESULT 3
US-10-127-032-16
; Sequence 16, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Bangera, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; TITLE OF INVENTION: BIOFILM FORMATION

```

; FILE REFERENCE: UIZ-070CP
; CURRENT APPLICATION NUMBER: US/10/127,032
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,190
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/344,142
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-127-032-16

Alignment Scores:
Pred. No.: 5,97e-45 Length: 1389
Score: 527.50 Matches: 151
Percent Similarity: 45.84% Conservative: 86
Best Local Similarity: 29.21% Mismatches: 205
Query Match: 20.40% Indels: 75
DB: 15 Gaps: 16

US-10-009-823A-1 (1-502) x US-10-127-032-16 (1-1389)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 4 AGTTTCAACATCGGCTCGAGCGGCATCCAGCGGCCTCTAGCGGCCTGAACGTCAACCGGC 63
QY 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
Db 64 AACACATCGCAACGCGCGCACGTCAGCTTCAAGCAATCCGCGCGGAGTTGCGCGAC 123
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 124 GTCTACGCGCGCTCGGTG---CTGGGTTCG-----GGCAACAACCGCGCGGC 168
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 169 AGCGCGTGTCTCGAGCTCTCGAGATGTTCAAGCAGGCGCAACATCGACTCGACC 228
QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
Db 229 AACAGCGTGTCTCGAGCTCTCGAGATGTTCAAGCAGGCGCAACATCGACTCGACC 288
QY 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
Db 289 GCGATCAGCTACCGCGCGCGGTCTCTTCAATCCGACAGCAGGAGTTTCACTGTCGAC 348
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
Db 349 AACACGCGTACCGCGCTCGAGGTATGCTCGTGGCGGCAAGCGCCAGTTGCGAAGCGC 408
QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
Db 409 GTGGTACCGACCTCAAGGTGAGCGGCGCCATCAGGCGCGCGAGGCCACTCGAGATC 468
QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
Db 469 CAGCAGTGTCAACACCTCAACTCGACGCTGAAG-----CCG 504
QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
Db 505 CCGACCGTGTACCGGCTC-----GATCCG 528
QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
Db 529 TCCGACGCGGTACTACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 588
QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
Db 589 TCCACACCATGAGCCAGTTCTTTCATCAAGAACGAGCGCGGACCGGATGCGACCGCGCG 648

```

RESULT 4

```

US-10-282-122A-14579
; Sequence 14579, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14579
LENGTH: 1239
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-14579

Alignment Scores:
Pred. No.: 7,87e-43 Length: 1239
Score: 506.50 Matches: 137
Percent Similarity: 43.34% Conservative: 81
Best Local Similarity: 27.24% Mismatches: 180
Query Match: 19.59% Indels: 105
DB: 16 Gaps: 11
US-10-009-823A-1 (1-502) x US-10-282-122A-14579 (1-1239)

Qy	8	GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla	27
Db	16	GTGTTGACGGTTTGGCGGGCGGTTCAGCGACCTCGACGTATCGGCAACATCGCG	75
Qy	28	AsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln	47
Db	76	AACGGAAACACGGTGGGCTTCAAGCGGACGACCGCGCAGTTTCCGACATGATCGGAAT	135
Qy	48	AspLeuAlaIleGlySerThrGlySerGlnGlnProAsnGlnAlaGlyMetGlyAlaGln	67
Db	136	TCGGTCGCA-----TCGGCGGTCAACAACTCGATCGGCATCGGCACGATG	180
Qy	68	ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr	87
Db	181	CTCGCATCGGTGCAGCAGCAGTTTCAGCGAGGCGACGATCACTCGAGCACGTCTGCTG	240
Qy	88	AspLeuAlaIleGlyGlyPhePheGlnValThrLeuGluAspLysValHisTyr	107
Db	241	AACGTTCGGATCAACGGCAACGGTCTTTCAGATGTCGAACAAACCGCGGTGACCACTAC	300
Qy	108	ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe	127
Db	301	TCGCGGACGCGACGTTCCAGCGGACAGAACGGCTACATCGTCAATTCGCGGGCGCTG	360
Qy	128	ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro	147
Db	361	AACGTGATGGGTACGCGCGAAGCGGACGCGGTGATCAACACCGCGGGCGACCGTGC	420
Qy	148	IleGlnLeuAspPheAsnAspThrValAlaLysSerProAlaLysThrSerThrAla	167

Db	421	CTGCAGGCG-----CCGACGACCAACATCGCGCCACCGCGGACG---ACCAG	465
Qy	168	LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro	187
Db	466	ATCACCGGCGGCTTCACTG---AACTCGAGGACGCGGTCGCGCG---ACGACGCG	519
Qy	188	TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn	207
Db	520	TTC-----AACTACACCGATCCGACGAGC-----	543
Qy	208	TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr	227
Db	544	TACAACTACAGCAGCTCGGTGCGAGGTTCGACACGCTCGCGGCTCGCAGAACGTGAAC	603
Qy	228	ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla	247
Db	604	CTGTACTTCTGAGAGCGCGACGAGCGCGGAGCGGTATGCGGGCGCGCGCGCGC	663
Qy	248	MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet	267
Db	664	AAGACGCGGACCGATCTCGGCTCG-----	687
Qy	268	SerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrPro	287
Db	688	-----GTCAAAGTTTCAGCACCGCGCGGACGATC	714
Qy	288	ThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeu	307
Db	715	ACGGGACGTCGACGCGC-----	732
Qy	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly	327
Db	732	-----	732
Qy	328	IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr	347
Db	733	-----CCGCGCGTGCAGC-----ACG	750
Qy	348	AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn-----	364
Db	751	AACGTGCGGCGGTTCTGTTCTCGATCCGACGACGCGGCGCGGACCGCGCAGAAC	810
Qy	365	-----SerThrAlaArgAsnGlySerSerSerThrArgArgTyr	377
Db	811	CTGACGCTCGACCTCACGGGCGACGACGAGTACGCGCGGCAAGACGGCATCAACATCTC	870
Qy	378	SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLys	397
Db	871	GCGCAGGACGGCTTCGAGCGCGGTCTGACGACGCTTCTCGATCGGCGGACGCGCAG	930
Qy	398	LeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArg	417
Db	931	GTACCGGCAACTACTCGAACGCGCAGACCTCCACGCTCGGCGCAGATCGTCTCGCAAC	990
Qy	418	PheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAsp	437
Db	991	TTCAACAAACCCGAGCGCTCGTGAACGTCGCAACAAACAGTACGTGAGTTCGCGCGG	1050
Qy	438	SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsn	457
Db	1051	TCGGGCGTCCGACAGATCTCCGCGCGCGGACGACGACGACGCGCGCTCGAGGCGAGC	1110
Qy	458	GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGln	477
Db	1111	GCGCTCGAAGAACTCGAACGTCGACCTGACGACCGAGCTGCTGAAGCTGATCACCGCG	1170
Qy	478	ArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAla	497
Db	1171	CGCAATACCGGCGGACGCGGACGATCAAGACGCGACGACCGTCCGACCGAGCTC	1230
Qy	498	LeuGluLeu 500	
Db	1231	ATCAACCTG 1239	

325 AspPheGly-----lleYsserGlnGln 332
Db CAATACGGCATGTTGAAATCAACAATAAAGATAATAAATATGCTCATTAAGAAATATCTC 1305
333 AsnMetTrp----- 335
Db AATATCTTTGAGTGGGTATCTTTCAGACAGCGTAAACAAATGTTTGTAAAC 1365
336 -----AlaGlyAlaProAlaSerAlaAla 343
Db GCGATGAAGGCTTAATACGGCTTCTTTAATGAAGGGGAGGCTAGCTAGCTTCT 1425
344 -----AlaIleGlyThr--- 347
Db AAATTCACCCACGCTACCCATCGCAGCAAGCATGATGATAGACAGCTTTAGGCACTAAA 1485
347 ----- 347
1486 CACCCATGCGCATGAGTTTATAGAGTGGGGGCGGAATGGAATTTAGAGTGATC 1545
347 ----- 347
1546 GTGCTGAGCTGGGAATAGTAGGGGGGTACAGCGCTAGGCTAATGTGTTTGAAGGG 1605
348 -----AspIleGlyLysLeuProSerMetMetPro--- 357
Db GCGCGTTTGCAATTTAATAACGAGGATCGCTTCGGGGCATGAACCGCCCTCTTTTGCAA 1665
358 -----IleGlnThrSerSerGlyAsnSer 365
Db TTTGACCTAAATGCGGTGATGCCCCCAACGATCAATTTAGCTTTTGGTCTCTCA 1725
366 ThrAlaArgAsnGlySerSerThrArgArgTyrSer-----Gln 379
Db GGGAGCTTTGACGATTCACAAAGCGTGGATAAGATTTCTGAAACTTATGCGATTGAGCAA 1785
380 AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
Db AACGGCTATACGGCGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845
400 GlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThr 419
Db GGGCGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845
420 SerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db AACGATCGGGCTTACAGGCTTTGAGGAGGCAATGCTTTTCTCAACCGGAACTCTGGG 1965
440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db CAACCTTAATCGGTGCGCTAATACGGGGGTAGGGTTCATTTTCAGGATCCAACTA 2025
460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGly 479
Db GAGTCTAGTATGTTGGATTTTGAAGCGGAGTTTAAACGAATTTGATTTGTTTCAAGGGGG 2085
480 PheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
Db TTTCAAGCGAACTCTAAGCGGTAACCATCCGATCCGATCCATCCATCCATCCATCCATCCAT 2145
500 LeuLysArg 502
Db CTCAGCAA 2154
RESULT 6
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications

FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 10
LENGTH: 1163020
TYPE: DNA
ORGANISM: Listeria innocua
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(end)
OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10
Alignment Scores:
Pred. No.: 2,8e-36 Length: 1163020
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 16 Gaps: 11
US-10-009-823A-1 (1-502) x US-10-398-221-10 (1-1163020)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db ATGAATCAACATCATGTATATACAGCTATTTCTGGATGAATGGTTCCTCAACAGCATATCA 639597
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db GTGACATCAATTAATATTCCTCAATCGCAACAGCAGCGGTATATAAACAAGCCGCTGT 639597
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db TTCAATGATTTACTTTACCAAAACACAATGGATCTGTTCAGGGCGGACTTTATGCTGA 639717
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db ACAAAACCAATAGCTTCGGTTCGGTTCGAAATTTGGGCAATTTTAAACCGATTATACA 639777
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db GCAGTTCTCCGACATCACTCGCGAGAAACAAGATGAGCACTGCAAGGCGCGGCTTT 639837
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db TTCAATGCTGGCGATAACGCTGGTGGGAATATCGTTTACACAGACAGAGCTAGCTTGA 639897
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db GTATCTGCACTAATCACTATTTTAAACAGCAACAGGAAATACGTTTGGGATATCAACA 639957
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db GATAAAATATGCAACGCTTTTAAACCGAAACTTGCAACCAATTCAAATCCCAATTAATAGC 640017
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db GCAATTCAGGCGAGCAACAAAAAATGATGATTAAGGGTAAACATTCCTCCATCTGATGG 640077
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db GCGGAAAAA---GATACGATTTCTCCGAG----- 640104
QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 640105 -----CTT 640107

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
Db 640108 TCTGTATACGTAATGCTGGTGGAAACATATAAATTCAGTCAATATGAAGCTGGAC 640167
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 640168 CCAGATGGGAGTGGCAATGTTCTCTATGAATGAATCAATG-----GAC 640215
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 640216 GGAAGCGTTAACTCCTCCA-----GTAACAGGACACCTTAATATC 640257
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 640258 AACGGCAAGGCGAATCTTACAAACCTGATGCA----- 640290
QY 294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 640291 ---CITAAACAACTCAATCAATCCACTGTTAACGGC----- 640326
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 640327 -----AAACAAGTCAAT 640338
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 640339 ATG-----GGCTTAAACCTTAAGTGGCTTAACA 640365
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSerSer 373
Db 640366 RACTAC-----GGTACCACCAA 640383
QY 374 ThrArgA-GTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 640384 GTATTCTCACCAACTCTCGACGGAAAGCGCTGCAACTGTAAAGAGACTACGCAAGTTACC 640443
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 640444 GATTCGTGTATATGCAAGTACGATTCTACACGGTACAGTTATCCAGTGGCCCACTT 640503
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 640504 GGGTGTGCTACTTCTCCATGAAGACGGCTTAGTCAAAATGGGCAATGGCAATATGTT 640563
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 640564 CCAGGATTATCTCTGGAGATGCAAGTTTACGGCGTTGTCGCCAAACGGAGCTGGCGGA 640623
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 640624 ATTACGGCTCTTCAATTAGAGGCTCAACCGTAGATTATCCCGGAATTCGTTAACTTA 640683
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 640684 ATGACATACCAAGTGGTTTCCAGGCAATCAAAAGTATTCCTGTTGCGAGATGACGTG 640743
QY 494 LeuGlnLysAlaLeuGluLeu 500
Db 640744 ATGAACAAATGTGAAGTTG 640764

RESULT 7

US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome, and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04

; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058

Alignment Scores:
Pred. No.: 1,098-35 Length: 3011208
Score: 484.50 Matches: 177
Percent Similarity: 42.21% Conservative: 137
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 16 Gaps: 11

US-10-009-823a-1 (1-502) x US-10-398-221-2058 (1-3011208)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 729803 ATGAATCAAACATATGTATACAGCTATTCTGGGATGAATGCGTTCACAAAGCATTATCA 729862
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 729863 GTGACATCAATAATAATTATGCCAATCGCAACACGACGAGGTATATAAAACAAGCGTCTGT 729922
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 729923 TTCAATGATTACTTATCCAAACACAATGGGATCGTTCGAGCGGACTTTATGCTGCA 729982
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 729983 ACACACCCCAATAGCTTCGGTTCCGGTTCCGAAATTTGGGCAATTTTAAACCGATTATACA 730042
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db 730043 GCAGGTTCTCCACATCACTGGCAGAAACAAGATGCAGCACTGCAAGCGCGGCTTT 730102
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 730103 TTCAATGCTGGCGATAAAGCTGCTGGGAATATCGTTTACACACGACGAGCTAGCTTTGCA 730162
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgile 134
Db 730163 GTATCTGACAAATACTATTTAACACGCAACAGGAAATACGTTATGCGGATATGCAACA 730222
QY 135 SerAsnAsnProAsnIleLysLysGlyThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 730223 GATAAAATGGCAACGTTTAAACGGAAACTTGCAACCAATTCAAATCCCATTAATAGC 730282
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 730283 GCAATTCACGGCAAGCAACAAAAAATGCTAGCTTAAGCGGTAACTTCCCACTTGATGG 730342
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSer 194
Db 730343 GCGGAAAAA---GATACGATTCTTCCGAG----- 730369
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 730370 -----CTT 730372
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
Db 730373 TCTGTATACGATAATGCTGGTGGAAAAACATAAACTTCAACTCAATATGAAAGCTGCGACA 730432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 730433 CCAGATGCGAGTGGCAATGTTTCTCTATGAATATGAATATCAATG-----GAC 730480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273

Db 5487 AATGCGCAGGGGAATTAAACAAACCAACGACGCA----- 5519
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 5520 ---CITAAAAATATCAAAATCAATTCACAGTAACGGC----- 5555
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 5556 -----AAACAAGTCAAT 5567
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 5568 ATG-----GGCTTAACCTTAAGTGGCTTAACC 5594
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
 Db 5595 AACTAC-----GGAAACAACCA 5612
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 5613 GTATTCTCACCACACTTCGACGGTAAGCGCTCGGACTGAAAGATTATGCGAGTTACC 5672
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 5673 GATTCTCGCTATATGCGAGTCAGTTACTCAGATGCTACAGTTATCCCGACTTGCCCAACTT 5732
 QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
 Db 5733 GCGGTGGCTACTTCTCAATGAAGACGGCTTAGTCAAAATGGGGAACGCGAATATGTT 5792
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 Db 5793 CCAGGATTATCTTCGGCGATGACAGTACGCGCTGTCGCCAAATGGCGTGGCGGA 5852
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 5853 ATTACGGGTCTCTCTAGAGAGTTCAACAGCTACACTTGTCCCGTGAATTCGTTAACTTA 5912
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 Db 5913 ATGACATACCAAAATGGTTCCAGGGCATACAAAGATTATTCGTGTCGGCAACGCTG 5972
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 Db 5973 ATGAATCAAACTGTGCACTTG 5993

RESULT 9

; Sequence 1, Application US/09790988
 ; Patent No. US20020127687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHIGENOBU, SHUJI
 ; APPLICANT: WATANABE, HIDEMI
 ; APPLICANT: HATTORI, MASAHIRA
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
 ; FILE REFERENCE: 081356/0159
 ; CURRENT APPLICATION NUMBER: US/09790,988
 ; PRIOR FILING DATE: 2001-02-23
 ; PRIOR APPLICATION NUMBER: JP2000-107160
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 640681
 ; TYPE: DNA
 ; ORGANISM: Buchnera sp.
 US-09-790-988-1

Alignment Scores:

Pred. No.: 2,4e-29 Length: 640681
 Score: 414.50 Matches: 132
 Percent Similarity: 39.96% Conservative: 73

Best Local Similarity: 25.73% Mismatches: 197
 Query Match: 16.03% Indels: 121
 DB: 9 Gaps: 13
 US-10-009-823A-1 (1-502) x US-09-790-988-1 (1-640681)
 QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 Db 372153 TCAAAATATGATGCTATAGTGGCTTACTAGCAAAATAATGATTACATGAAATATATCC 372212
 QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
 Db 372213 AATAATATTGCTAAGCATCAACTATAGATATAAATCTCGTPAAACCTCTTTTTCAT 372272
 QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 Db 372273 ATGTTTCTCATTCTTTTATTCAAAATACTACTAAT-----GGA 372311
 QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 Db 372312 TACGGAGTGGGTATTTCAGTATTATATACAAACTTTAATAATGGCATGTTAGTCAAACT 372371
 QY 84 AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGlu 102
 Db 372372 GCACGAGATTTCGATTAGGAATTATAAAAGACGGCTTTTCGGTCTGTAGACAGATCAA 372431
 QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
 Db 372432 GGTCAATGCTATTATACAGAGATCGGCAATTCTTCCTCGATAAAGATCAAAATATTATC 372491
 QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136
 Db 372492 AATATTCAAGGTATGATTAATCACTTAATCATCTTCTCAAAAGTGATTTTAAT 372551
 QY 137 AsnProAsnIleLysGlyGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156
 Db 372552 AAT-----AGATCCAACTTAGAACCTATTATTTTA-----AAAAATTTCTAAT 372593
 QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGly--- 175
 Db 372594 ATTTTAAAAAACAACTTCTCGAATCATGTATAAAGCGTTTGGAAATCGTAATACT 372653
 QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189
 Db 372654 GAATCAAAAAGCAGTGTGGATAATTCTGACAACTATCTAAACAGAGATATACATG 372713
 QY 190 AlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSer 209
 Db 372714 ACTTATATTAGC----- 372725
 QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229
 Db 372726 -----ATATATAATAAAGAGGAGGAAAAAGAGATATTACTGTCTCT 372767
 QY 230 PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn 249
 Db 372768 TTTAAT-----AAAAAGAAACAAATAAATGGACAGTAAT-----GTGGAATCAAT 372815
 QY 250 ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly 269
 Db 372816 GATTCTGATGATAAAGAGACTATAAAAAAT-----AGTTTC 372851
 QY 270 ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly 289
 Db 372852 GATTTAACGTTTAATGATGATGCGCAATTAACCTCTGATATGTTTTTATATTACATCT 372911
 QY 290 SerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGln 309
 Db 372912 AAAGATTCTAAAAAG----- 372926
 QY 310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLys 329
 Db 372927 -----TATGAAATATCACTTTAAATTTA----- 372950

QY 330 SerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349
Db 372950 ----- 372950
QY 350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
Db 372951 -----ACAGGTACTATAGAACAAATCAATCAAT 372974
QY 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
Db 372975 TCTGATGTTCTTGGGAAGAACATCTCAAAACGGATACCCCTCAAGGTAAATTTAAACACA 373034
QY 390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
Db 373035 TTTGATATTCTTACTAATGTTGAATTTTGGACATATTGCAATCAAAACACACACACA 373094
QY 410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsn 429
Db 373095 ATAGGTCAAAATATTATCAAAATTTATCAATCCAGAAAAATTAACAACCTGAAGTGT 373154
QY 430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu--ProGlyThr 448
Db 373155 AATTATGGTCTGCTCTGAGATCTAGGTGAAGCAAAACACACAAATGAAGCGGTATT 373214
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Db 373215 CAAGAATCAGAGGTGTTAAGCAATAAAACGCTAGAGTATCAAAATGTTGATTTGAATAAA 373274
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db 373275 GAATTAATCAATATGATTATAGCACACGTAATTAATCAATCTAACGCTCAATCTTTTAA 373334
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 373335 ACAGAATAAAATAATTAATACATTAATAATAATTACAG 373373

RESULT 10
US-10-335-977-56
Sequence 56, Application US/10335977
Publication No. US20040052799A1

GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335.977
FILING DATE: 30-Dec-2002
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:
LENGTH: 666 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...666
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-10-335-977-56

Alignment Scores:
Pred. No.: 2,99e-21 Length: 666
Score: 296.00 Matches: 78
Percent Similarity: 48.94% Conservative: 37
Best Local Similarity: 33.19% Mismatches: 98
Query Match: 11.45% Indels: 22
DB: 16 Gaps: 3

US-10-009-823A-1 (1-502) x US-10-335-977-56 (1-666)
QY 272 ThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAla 291
Db 13 TCTATAGAGGTGAGGGGGCGGAATTTTAGATGATCGTCGCTGAGCCTGGGAA 72

QY 292 ThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSer 311
Db 73 TTAGTAGGGGGGTACGGCTAGGCTAATGTGTTGAAGGGGGCGGTTGCTATTTAAT 132

QY 312 AlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGln 331
Db 133 AACGACGGATCGCTTTCGGGCGATGAACCCGCTCTTTTTCGAATTTGACCCCTAAAAAT 189

QY 332 GlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLys 351
Db 190 -----GGCGCTGATGCC----- 201

QY 352 LeuProSerMetMetProIleGlnThrSerSerGlyAsnSer-----ThrAla 367
Db 202 ---CCCCAACGCATCAATTTAGCTTTTGGTTCTCCTCAGGGAGCTTTGACGGATTGACAAGC 258

QY 369 ArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu 387
Db 259 GTGGATAAGATTCTGAAACATTATCGGATTAGCAAAACGGCTATCAACGGGGCGATTG 318

QY 388 ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal 407
Db 319 ATGGATGTCGGCTTTGATTCGATGGGTGCTTTTAGGGGGCGTTTCAGTAATGGCAGGACT 378

QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 379 TTAGCGCTCGCTCAAGTGGCTTTAGCGAATTTTCGCTAACGATGCGGGCTTACAGCGTTG 438

QY 428 GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447
Db 439 GGAGGCAATGCTTTTCTCAAAACCGGAACTCTGGCAAGCCTTAATCGTGGCGCTAAT 498

QY 448 ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer 467
Db 499 ACGGGGCGTAGGGGTTCAATTTTCAAGATCCAAATAGAGTCTAGTAATGTGGATTGAGC 558

QY 468 ArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerVal 487
Db 559 CGAGGTTTAAACGAATTTGATTTGTTCAAGGGGGGTTTCAAGCGACTCTTAAGCGGTA 618

QY 488 ThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 619 ACCACATCCGATCAATCTTAAACCCCTATTGAATCTCAAGCAA 663

RESULT 11

US-10-335-977-1
 ; Sequence 1, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: Windows NT 4.0
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/335,977
 FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/593,002
 FILING DATE: 17-Dec-1997

ATTORNEY/AGENT INFORMATION:
 NAME: Mandragoras, Amy E.
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: GTN-018
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 789 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Helicobacter pylori

FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...789
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-335-977-1

Alignment Scores:
 Pred. No.: 2,61e-20 Length: 789
 Score: 288.00 Matches: 107
 Percent Similarity: 28.68% Conservative: 39
 Best Local Similarity: 21.02% Mismatches: 109
 Query Match: 11.14% Indels: 254
 DB: 16 Gaps: 9

US-10-009-823A-1 (1-502) x US-10-335-977-1 (1-789)

Qy 1 MetVetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 1 ATGCTCGCTCTCTATAGTGGCACTTCAGGATGCTGCCCAACAAACGACATTGAC 60
 Qy 21 ThrValSerAsnAsnLeuAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 61 ACCACTTCAACACACATCGCATGTCATACACCGGTTTAAAAAATCTCGCGCGAT 120
 Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 Db 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60

Db 121 TTTAACGACTTGTGTTTACCAAGCGATGCAATACGGCGCACCAACACACGACT 180
 Qy 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 Db 181 TTATCGCCAGATGGCATGGAAGTGGCTTAGCGGTACGCCCTAGTGGCATACCAAAATG 240
 Qy 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 Db 241 TTTTCGAAGCGACCCCTAAAGAAACGGAGATAATTTAGATATTGCTATTACAGGTAAA 300
 Qy 95 GlyPheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 Db 301 GGCTTTTTCAGTTCAGCTTCTGATGGCACTACCGCTTACCAAGGAGCGGAAATTC 360
 Qy 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 Db 361 AAGCTAGACGAGCAGGCAATCTTGTAAACAGCGAGGCTATCTCTCATC----- 411
 Qy 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db 412 -----CCTCAATC-----ACTTTA----- 426
 Qy 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 Db 427 -----CCCGAAGACACACGCAA-----GTGAAT 450
 Qy 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
 Db 451 ATCGGT----- 456
 Qy 194 SerTyrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 Db 456 ----- 456
 Qy 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 Db 457 -----GTGGATGGCAGC 468
 Qy 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 Db 469 GTGACGCTGACT----- 480
 Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 480 ----- 480
 Qy 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 Db 480 ----- 480
 Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 480 ----- 480
 Qy 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 480 ----- 480
 Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 480 ----- 480
 Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 480 ----- 480
 Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 480 ----- 480
 Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 481 -----CAAGGCTTGCAACACGACTTCTAACGCTGATCGGG-----CAATC 519

314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
507 -----
334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
507 -----
354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
507 -----
374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValThrIleThr 393
507 -----
394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
508 -----CAAGGCTTGCAACGACTTCTACGTCATCGGG-----CAAAATC 546
414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
547 ACTTGGCTAATTTGTCAATCCGCGGGGCTTCATTCTATGGGGATAATTTGTTTCC 606
434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
607 ATCAACCAACGCTAGCGCGATGCGATTGTGGGCAACCGGATTCTCAAGGCTTAGGCAAG 666
454 LeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMet 473
667 TTAGGCAAGCTTTTGGAGCTTAGTAACGTAGATGTGTAGAGAAATGACAGATCTA 726
474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
727 ATCAACCGCTCAAGGCTTATGAAGCAATTTCTAAAGCATTCAAACCGCTGATGCCATG 786
494 LeuGlnLysAlaLeuGluLeuLysArg 502
787 CTCCAACAGTCAATTCCTCAACGC 813

RESULT 13

US-10-335-977-3
; Sequence 3, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...837
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-335-977-3

Alignment Scores:

Pred. No.:	2,848-20	Length:	837
Score:	288.00	Matches:	107
Percent Similarity:	28.68%	Conservative:	39
Best Local Similarity:	21.02%	Mismatches:	109
Query Match:	11.14%	Indels:	254
DB:	16	Gaps:	9

US-10-009-823A-1 (1-502) x US-10-335-977-3 (1-837)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
DB	52	ATGTCCTCGCTCTCTATAGTGCACCTTCAGGATGTCGCCCAACAAACGACATTCAC	111
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
DB	112	ACCACTTCAACAAACATCGCAATGTCAATACCAACCGGGTTTAAAAATCTCGCGCGAT	171
QY	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
DB	172	TTTAAAGACTTGTTTTACCAAGCGATCAATACCGCGCACCAACACGACGACT	231
QY	61	-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle	74
DB	232	TTATCGCCAGATGCGATGGAAGTCGGCTTGGCGCTAGTGGATTTACCAAAATG	291
QY	75	PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys	94
DB	292	TTTTCGCAAGCGACCCCTTAAGAAACGAGAAATATTAGATATTGTTATACAGTAA	351
QY	95	GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsn	113
DB	352	GGCTTTTTCAGTCCAGCTTCTGTGATGGCACTACCGCTTACACAGGAGCGGAATTC	411
QY	114	ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg	133
DB	412	AAGTAGACGACGAGGCAATCTGTAAACGAGCGGGCTATCTCTCTATC	462
QY	134	IleSerAsnAsnProAsnIleLysLysGluThrIleGluProIleGlnLeuAspPheAsn	153
DB	463	-----CCTCAATC-----ACTTAA	477
QY	154	AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn	173
DB	478	-----CCGAAACACACACGCA	501
QY	174	LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu	193
DB	502	ATCGGT-----	507
QY	194	SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro	213
DB	507	-----	507

QY	214	MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla	233
DB	508	-----GTGGATGGCAGC	519
QY	234	ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp	253
DB	520	GTGAGCGTGACT	531
QY	254	GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe	273
DB	531	-----	531
QY	274	SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys	293
DB	531	-----	531
QY	294	AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn	313
DB	531	-----	531
QY	314	PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyLeuLysSerGlnGlnAsn	333
DB	531	-----	531
QY	334	MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	353
DB	531	-----	531
QY	354	SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer	373
DB	531	-----	531
QY	374	ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr	393
DB	531	-----	531
QY	394	SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle	413
DB	532	-----CAAGCCTTGCAACGACTTCTAACGTCATCGGG-----CAATC	570
QY	414	ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyLysAsnAsnHisTyrSer	433
DB	571	ACTTTGGCTAATTTTGCTCAATCCGCGGGCTTCATTCATGGGGGATAATTGTTTTC	630
QY	434	AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys	453
DB	631	ATCACCACCTAGCGCGATCGGATTGGGCAACCCGGATTCTCAAGGCTTAGGCAAG	690
QY	454	LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet	473
DB	691	TTAAGGCAAGCGCTTTTGGAGCTTAGTAACGTGAGATTGGTAGAAGAAATGACAGACTA	750
QY	474	IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet	493
DB	751	ATCACCGCTCAAGGGCTTATGAAGCCAAATTTTAAAGCATTCAAAACCGCTGATGCCATG	810
QY	494	LeuGlnLysAlaLeuGluLeuLysArg	502
DB	811	CTCCAAACAGTCAATTCCTCCATCAACGC	837

RESULT 14
 US-09-974-300-1888
 ; Sequence 1888, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berk, Randy M.
 ; APPLICANT: Clausen, Ib Groth
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; TITLE OF INVENTION: Expression
 ; FILE REFERENCE: 10085.500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; CURRENT FILING DATE: 2001-10-05

;; PRIOR FILING DATE: 2001-03-27

;; NUMBER OF SEQ ID NOS: 8481

;; SOFTWARE: FASTSEQ for Windows Version 4.0

;; SEQ ID NO 1846

;; LENGTH: 1524

;; TYPE: DNA

;; ORGANISM: Bacillus licheniformis

US-09-974-300-1846

Alignment Scores:

Align. No.: 2,46e-12 Length: 1524

Score: 215.50 Matches: 130

Percent Similarity: 37.2% Conservative: 81

Best Local Similarity: 22.93% Mismatches: 207

Query Match: 8.33% Indels: 149

DB: Gaps: 28

US-10-009-823A-1 (1-502) x US-09-974-300-1846 (1-1524)

```

QY 5 LeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
DB 25 CTGGAATTCGCAAAAGAGGTTGTCGCGCAGCAATCAGCTCTGAGCGTGACAGCAAC 84
QY 25 AsnIleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValValPheGln----- 42
DB 85 AATATATCCACGCCAATACCGAGGGTATTCAAGACAGCGTGTATCATTTAAGCATCA 144
QY 43 -----AspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyPro 59
DB 145 ACCCTTATCCGTCGTCGATCAAGAGAC---TCTGTCGACTTCGAGGC----- 189
QY 60 AsnGlnAlaGlyMetGlyAlaGlnValGlySerValargThrIlePheThrGlnGlyAla 79
DB 190 ---CAGATGGGTACAGCGGTGAATTCGTTTCGTTGAAGGGTT-----AGGACAGC 240
QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnVal 99
DB 241 TTT-----LeuAsnAsp-----CTG 246
QY 100 ThrLeuGluAspLysValHisTyrArgAlaGlyAsnPheargPheThrGlnAspGly 119
DB 247 GACTATCAATACAGCGCAGAGATACAAAGCGCGCTATTACAAATGCAAAAGTGGAGCT 306
QY 120 Phe-----LeuAsnAsp-----123
DB 307 TTTAACCAAAATGAAGGATTGATGAATGAACGACCGGACTAAACAAAGTCTG 366
QY 124 -----ProSerGlyPhe 127
DB 367 AATTCATCTGGAATTCATTGAGGAATTCAGCAACAAACCCGACGCGCTGCT 426
QY 128 ThrLeuMet-----GlySerArgIleSerAsnAsnProAsnIleLysLysGluThr 144
DB 427 TCTCTTGTGGCAAAAGGCGCGGTGCGCGATACATTTAACACTCTGTATGATCG 486
QY 145 LeuGluProIle-----GlnLeuAspPheAsnAspProThrValAla 158
DB 487 CTGCAACACCGTTCAAATCAAAATTTAGGCGCAGAGATTGATCAAAATGTGATGAGATT--- 543
QY 159 LysSerProAlaLysThrSerThrAlaLeuAsnAlaVal-----172
DB 544 AACTCGGTTTTCGCGAGTGAACAGCAATTCGCAACAAATCGCAACGCGCAAC 603
QY 173 -----AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
DB 604 GGCTATCTCCGCAACGATTTGTACGATAGCGGTGATCAGCTCTTGTGATGAATCTTCA 663
QY 188 TyrPheAlaLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207
DB 664 ATGGTTAATCAAGTCACTAGCAAGAAACCGCGGCAATCCACTCCGCAACCGCTGAA 723
QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227

```

```

DB 724 GGCATTGTGTCA-----ATAGAAGTCTTTGACGTAAACGGCCAGTCT-----CTGGGC 771
QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
DB 772 AAGTTCCTTGACGGGCCAAACATTCACGACTGAACCT-----GTCAAA 813
QY 248 MetAsnProSerGluAsp-----GlySerAlaAlaSerGlyThrAsp--- 261
DB 814 GTGAATTACGACAAACGATACAGGCTTGTGACGGGGTTTCCCTCGCGGGAGCGACATC 873
QY 262 -----SerAlaGlyLeuLeuMetSer-----GlyThrMetThrPhe 273
DB 874 GGCATCGATTGCTTTACTAGCAAAAGGTCCTGCTCGGTTTAATCGAATCTTTACGGAT 933
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 934 ATGTCAAACGGTGAAGAAAGGTTTG-----TATCCGAAATGCTCCGATCTG 984
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 985 GATACGATGCGCTTCTCTTTTCGAAAGAGTTTAATGCATCCATCAA-----AGCGGG 1038
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 1039 TTTACTTCTATCGGGA---GAACCGGGAGCATTTCTTTTGATTTTACCGGAGGAG--- 1092
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 1093 -----GCTGAGCCTGCAAGGAGCGGCC-----GGCAAAATCAAA 1128
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg-----AsnGlySer 371
DB 1129 GTAGCCGATGATCATGATTTCAAAAGCGCATAAAGTCGCGCATCGCTGAATGGTGAA 1188
QY 372 SerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThr 391
DB 1189 GCGAGC-----GACAAACGCCAAGCGACAAATCTCGCTAATGCTCTTC 1230
QY 392 IleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyr 411
DB 1231 ACCAAG-----AAATTCNAATCGGTGACAAAACGACGACCGCTTTTACACTATTAC 1281
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHis 431
DB 1282 -----GCAGGATCATCGTGAATGGCGTAAAGCTCAGGAACAAACAGA 1329
QY 432 TyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyr 451
DB 1330 CTGCGCAAAATACAGAAACG-----CTTGTAACACACTGCTGAATTG 1371
QY 452 GlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetVal 471
DB 1372 AACGACAGTCTGTT-----AGCGCGTATCATTTAGCAGAAATGTCTG 1416
QY 472 AsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAsp 491
DB 1417 AACATGATCCAAATTCGAGCATGCTTATAACGCGCTCGGAGATGATTACGCTGCAGGAC 1476
QY 492 ThrMetLeuGlnLysAlaLeu 498
DB 1477 GAAGTCTCTGATAAGATTATT 1497

```

RESULT 16

US-10-335-977-78

; Sequence: 78, Application US/10335977

; Publication No. US20040052799A1

; GENERAL INFORMATION:

; APPLICANT: DOUGLAS SMITH et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

; RELATING TO HELICOBACTER PYLORI FOR

; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031

; CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 459 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...459
SEQUENCE DESCRIPTION: SEQ ID NO: 78:

US-10-335-977-78

Alignment Scores:
Pred. No.: 1,27e-11 Length: 459
Score: 201.50 Matches: 47
Percent Similarity: 53.68% Conservative: 26
Best Local Similarity: 34.56% Mismatches: 52
Query Match: 7.73% Indels: 11
DB: 16 Gaps: 3

US-10-009-823A-1 (1-502) x US-10-335-977-78 (1-459)

Qy 378 SerGlnAspGlyTyrProGlnGlyAspLeuValThrIleThrSerGluGlyLys 397
Db 49 AGTCAAGCGCAAGCCTAGGGGCAATTTTAGACACATGCCATTGAAGAAATGGGGTG 108
Qy 398 LeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArg 417
Db 109 ATTCTCTAGCCTTTAGTAACGGGGTGTAGAGCGGTCGCTCGCATCGGTATTTAGCT 168
Qy 418 PheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr-----SerAla 434
Db 169 TTCACACAGATCAAGGCTTAGGAAATCCGCGGTAACTCTATGAATGCAAGAGGC 228
Qy 435 ThrLeuAsp-----SerGlyGlyProGluPheGlyLeu-----Pro 446
Db 229 ACCATTATGCGAAACACAGACCCCTAAGCGGTAAACCCATTTTAGGGTGGGACGAAGAG 288
Qy 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnIleGluThrSerAsnValAspMet 466
Db 289 GGCACAGCTCAAGTTGGGAAATACGCACAAATATTAGAAACGACGACGTAATGCC 348
Qy 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486

Db 349 GGGNACGCCCTAACCAATCTCATTTTAATGCAAGAGGCTATTCTATGACGCTAGAGCC 408
Qy 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuLeuLysArg 502
Db 409 TTTGGCGGGCGGATGATCATGATCAAGAGCCATTAGCTTGAAGAAA 456
RESULT 17
US-10-210-296-2
; Sequence 2, Application US/10210296
; Publication No. US20030021802A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED
; TITLE OF INVENTION: METHODS
; TITLE OF INVENTION: AND MATERIALS
; FILE REFERENCE: PCI0589A
; CURRENT APPLICATION NUMBER: US/10/210,296
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/689,065
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 5445
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-10-210-296-2

Alignment Scores:
Pred. No.: 3,39e-09 Length: 5445
Score: 193.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.46% Indels: 0
DB: 14 Gaps: 0

US-10-009-823A-1 (1-502) x US-10-210-296-2 (1-5445)

Qy 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
Db 3 AACGTAGACATGAGCAGAGAAATGGTTAATATGATTATTATTCACGGTGTTCAGATG 62
Qy 483 AsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLysArg 502
Db 63 AATAGTAATCTGTTACACAGACACACATGCTACAAAGACACTTGAACCTAAAGCGT 122

RESULT 18

US-10-449-462-2
; Sequence 2, Application US/10449462
; Publication No. US20030202983A1
; GENERAL INFORMATION:

; APPLICANT: Pfizer Products, Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS AND RELATED METHODS AND
; TITLE OF INVENTION: MATERIALS
; FILE REFERENCE: 3153.00187/PCI0589A
; CURRENT APPLICATION NUMBER: US/10/449,462
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US/09/689,065
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US Prov. 60/160,922
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US Prov. 60/163,858
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 5445
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-10-449-462-2

Alignment Scores:
Pred. No.: 3,39e-09 Length: 5445

Score: 193.00 Matches: 40
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.46% Indels: 0
 DB: 15 Gaps: 0

US-10-009-823A-1 (1-502) x US-10-449-462-2 (1-5445)

QY 463 AsnValAsnMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
 Db 3 AACGTAGACATGAGCAGAGAATGTTAATATGATTATTCACGCGGTTTTCAGATG 62
 QY 483 AsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 Db 63 AATAGTAATCTGTTACACAGACACACAACTGCTACAAAGACACTTGAACCTAAGCGT 122

RESULT 19

US-10-335-977-82
 ; Sequence 82, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: Windows NT 4.0
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/335,977
 FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
 REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 1395 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...1395

SEQUENCE DESCRIPTION: SEQ ID NO: 82:

US-10-335-977-82

Alignment Scores:

Pred. No.: 1.12e-09 Length: 1395
 Score: 189.50 Matches: 59
 Percent Similarity: 43.14% Conservative: 29
 Best Local Similarity: 28.92% Mismatches: 81

Query Match:

DB: 7.33% Indels: 35
 16 Gaps: 6

US-10-009-823A-1 (1-502) x US-10-335-977-82 (1-1395)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHiserThrGlyLeuGly 20
 Db 1 ATGAACACACACCTTATTAACCGTTATAGCGGATCAAGACCCACAGTGTGTTATGAC 60
 QY 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 61 AGCCTTCCATAATATCGCAATGCTCAATACCTTTAGGCTATCGCTCAATATCGGAG 120
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyPro 60
 Db 121 TTATAGACCTTGTTCCTTCGCATTAGACGGCTTTGAACGCCCAATCCGTGTGCTGAAC 180

QY 61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGln---GlyAla 79
 Db 181 GACCGAAATTTACGGCGTT---ACAGGATCAGGCAATGCTTCTTAATAAAGACGGCGAA 237

QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnVal 99
 Db 238 TACATGCCCAAGTGAAGGGGAATTCACATGGCGTATCAGGGCAAGGTTGTTGTGATA 297

QY 100 -----ThrLeuGluAspLys 104
 Db 298 GGGCCCAATAAAACGGGGAATGACCATTAATAAAGATGCTTTAGCAAAAACAGGAT 357

QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124
 Db 358 AATTTCCTTAACCGCGTCCGCGCAATTTCCGCGAGAGCGCGATGCTATTATGTAACCCCT 417

QY 125 SerGlyPheThrLeuMetGly-----SerArg 133
 Db 418 GAGGCTATATGCTATGCTATGTTGATTGAAAATAATCAAGACGGCGCTCAATCC 477

QY 134 IleSerAsnAsnProAsnIleLys-----GluThrLeuGluProIleGlnLeu 150
 Db 478 ACCGCTAGAGATGAAGACATTAATAAATTCGAAACACCCCTTCGCCCTTACAAATC 537

QY 151 AspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 Db 538 -----CCCCAGATTGACTTACAGCGCGTGTCTAGCAGAAAGTGAATATC 585

QY 171 ValValAsnLeu 174

Db 586 AGCGTGAATCTA 597

RESULT 20

US-10-369-493-28400
 ; Sequence 28400, Application US/10369493
 ; Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B.

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 28400

LENGTH: 1965

TYPE: DNA

ORGANISM: Burkholderia fungorum

US-10-369-493-28400

Alignment Scores:

Pred. No.: 159e-08 Length: 1965
Score: 180.50 Matches: 144
Percent Similarity: 34.25% Conservative: 93
Best Local Similarity: 20.81% Mismatches: 224
Query Match: 6.98% Indels: 233
DB: 15 Gaps: 28

US-10-009-823A-1 (1-502) x US-10-369-493-28400 (1-1965)

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLeuThrHisSerThrGlyLeuGly 20
DB 1 ATGTCCTCAATCTCATCAATCTCGGCTCAGTGGACTGAACGCGACCCAGTGGGACTTCACG 60
QY 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrIleGlyGlnGlnValVal 40
DB 61 ACCACGGCCAGAAATCAGAACGGTTCGACCGCGGTACACCATCGAAACCGCGTA 120
QY 41 PheGlnAsp 43
DB 121 TATGCGGAAGCGGCGCCAGTACACGGGTTCGGGTATCTGCGCAAGGGGTCTCGACC 180
QY 44 -----LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGly 58
DB 181 ACACCGTGTACGCGCCAGTACAGCCAGTATCTGACCATGAGCTGAACACGCGCAGTGC 240
QY 59 ProAsnGlnAla 62
DB 241 TCACGAGTTCGTGTGCGGTACACACGCTGATCTCGCAACTGCAATCTGATCGGC 300
QY 63 GlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluPro 82
DB 301 AGCCCGACTCGCGGATTCGAGCGCGATCACCAGCTATTTCACCGCTTCGCAACAGTC 360
QY 83 GlyAsnSerValThrAspLeuAlaIle -----GlyGlyLysGlyPhe 96
DB 361 TCGAACACGCTCGACCTCGACGCGCCAGACCGCGATGAGCGGCGCAG ----- 414
QY 97 PheGlnValThrLeuGluAspLysValHieTyrThrArgAlaGlyAsnPheArgPheTh 116
DB 415 -----ACGTGTGTGAACCATCAACGCGCGGTGACGAATACAGCGCGTGC 465
QY 116 rGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAs 136
DB 466 CAGACGCTCAACACGCAACTCACCAACCGTTTCGCAGATTACAGCTACTCGCAGCAG 525
QY 136 nAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProTh 156
DB 526 ATCGCGCACTGACGACGACAGATCGCCAGCGCAGCACGC ----- 565
QY 156 rValAlaLysSerProAlaLysThrSer -----ThrAlaLeuAsnAlaValAlas 173
DB 566 -----AAGGCGACCGCGCAACAGTGTGTCGACCGCGCATCTC-GCCGTGTCCAA 617
QY 173 nLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
DB 618 CCTGTCCAGTTGATCGCGTGAACGTCGTCAACAGCAACGCGACGTACAGTGTTCAT 677
QY 193 uSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPr 213
DB 678 GAGC -----AATGGCCACGCGCTGTCGTCTCCACCAAC -----AGCTAC -----AA 719
QY 213 oMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAl 233
DB 720 CTTGGGCAACCGGCTTCCACCGGCGCACAGTGAATGTCTCGTGTGAGTATCTCGCCA 779
QY 233 aProSerSerThrGlySerLysThrPheGluTyrLeu ----- 245
DB 780 GCGCGCGCAACCGCGCGCGAGCGCCGCAAGACTGCGCGACAGCAAGATCGACGCGCG 839
QY 245 ----- 245
DB 840 CACGCTCGGCGCGCTGTGTTGCTTCGCGACGCAACGCTCGATCCGCGCGAGGCGCACT 899

QY 246 -----ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly ----- 259
DB 900 CGGCGCGATCGCCCTGAGCTTCTCGCGCAGGTCAACGCGCAGACGGGCTCGGTATCAC 959
QY 259 ----- 259
DB 960 GCTCGCGCGCCCAAGGCGCGCTGTTCTCGTGGGCGCGCCGACCGTCTACGCGAA 1019
QY 260 -----ThrAspSerAlaGlyLeuLeuMetSer ----- 268
DB 1020 CACACAGAACCCGCGACCGGTCTGTAAGTGTCTGTCGGGACGCCACGACGCGCGAC 1079
QY 269 -----GlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThr -----Al 284
DB 1080 CACCGCGACTACACGCTGGCTACAAACGCGCACCATCTACACGCTGACCGCAATTCGAC 1139
QY 284 aPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVa 304
DB 1140 GGGCACCGTGTGGTTCGGCACT -----AACCTGACCGACGCG -----AT 1181
QY 304 lAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLe 324
DB 1182 CAACGGCTG -----AATTTCTGACACACCGCG -----ACGAT 1214
QY 324 uAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaA 344
DB 1215 GAACGCGCGGATTCGTTTCACAGTCGACCGCAGC -CGCGCGCGCTGAACAGTTTCGCGA 1273
QY 344 lalleGlyThrAspIleGlyLysLeuProSerMetMetProIle -----G 359
DB 1274 CCGCACCAACGACGATCGCGGATCGCGCGCGCACCGCTGCTGCGCGCGGTGCG 1333
QY 359 lNTrSerSerGlyAsnSerThrAlaArgAsnGlySer ----- 371
DB 1334 CGAGCAACACCGCGCACCGCATACGAGGCGACGCTGACCGCGCGCTACACCATGC 1393
QY 372 --SerSerThrArgArgTyrSerGlnAsp -----GlyTyrProGlnGlyA 386
DB 1394 CGAATCTGACACACCGCTGTCTGACGAGCGCGGCTGTCCGCGCTTCCGCGCGGT 1453
QY 386 sPLeuValAspVal ----- 390
DB 1454 CGACGCTGACGCGCGGTACGCTCCACCAACGTCACGATCCACGCGCGCGCGACGG 1513
QY 391 -----ThrIleThrSerGluGlyLysLeuGlnGlyL 401
DB 1514 TGGTCCGCTATTCGTCGCGACCGCGCGCTCACGATCAACACGCGCGCGCGCG 1573
QY 401 yTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla -----ArgPheT 419
DB 1574 AGATGACAAATGTGTGCGTGACG --ATCAGCAACGCAACCGCTGCGCGCGCAAGTTCA 1630
QY 419 hr -----SerGluAspGlyLeuArgGluGlyAsnAsn -----HisT 432
DB 1631 CCATCGCGCGCAACACCGCGCGCACCAACGCGTGCACACGCGCTGGCATTTGCGAAT 1690
QY 432 yrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly ----- 447
DB 1691 TGTCCACCGCAAGCGATGTGCGCGGAGCGTCACTGACGCGCGGTATGCGAACT 1750
QY 448 --ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn ----- 463
DB 1751 AGTCAACAATATTCGCAAC --CAGACCAACCATCCAGACCTCGAGCGCGCGCAGA 1807
QY 464 -----ValAspMetSerA 468
DB 1808 GTTTCGCTGTGACGAGATCACACGCGCGCAGCTCGTTTCGGCGGTGACATCAACG 1867
QY 468 rGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValT 488
DB 1868 AAGACCGCACCACTGCTTTCAGTATCAGCAACTGTATCAGGCGCAACAGCAAGTCACTCC 1927

QY 488 hrThrAlaAspThrMetLeuGlnLysAlaLeu 498
Db 1928 AGACCGCGAGCCCTGTTCCAGACGATATC 1959

RESULT 21
US-10-369-493-31160
; Sequence 31160, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 473/74
; SEQ ID NO 31160
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31160

Alignment Scores:
Pred. No.: 1.59e-08 Length: 1965
Score: 180.50 Matches: 144
Percent Similarity: 34.25% Conservative: 93
Best Local Similarity: 20.81% Mismatches: 224
Query Match: 6.98% Indels: 233
DB: 15 Gaps: 28

US-10-009-823A-1 (1-502) x US-10-369-493-31160 (1-1965)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1 ATGTCAATCTCATCAATCTCGCTGCTAGTCACTGAGTCAACGAGCCAGTGGGACTCAGC 60

QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 61 ACACCGCGCCAGAACATCATGACGACGCTGACGCGCCGCTACACCATCGAAACGCGGTA 120

QY 41 PheGlnAsp----- 43
Db 121 TATCGGAAGCGCGCCAGTACACCGGTTCCGGCTATCTGCGGCAAGGGGTCTCGACC 180

QY 44 -----LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGly 58
Db 181 ACACCGGTGACCGCCAGTACAGCCAGTATCTGACCTAGCTGAGTGAACACGCGGAGTCG 240

QY 59 ProAsnGlnAla----- 62
Db 241 TCCAGGAGTTCGCTGTGCTAGCTACACACGCTGATCTGCGAATCTGATGATGATGCGC 300

QY 63 GlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluPro 82
Db 301 AGCCCGACTGCGGGATTGCGAGCGGATCACCAGCTATTTTCCCGGCTTGCAGAACGTC 360

QY 83 GlyAsnSerValThrAspLeuAlaIle-----GlyGlyLysGlyPhe 96
Db 361 TCGAACACGATCGAGCTCGGCGCGCCGACCGCGGATGAGCGCGCGCAG----- 414

QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArg-AlaGlyAsnPheArgPhe 116
Db 415 -----ACGCTGGTGAACAGATCAACGCGCGGTCAGCAATACGACGCGTTCGCG 465

QY 116 rGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAs 136
Db 466 CAGAGCGTCAACACGCAACTCACCACACCGGTTTCGACAGATTAAACGCTACTCGCAGCAG 525

QY 136 nAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProTh 156
Db 526 ATCGCGCACTGAACGAGCAGATCGCCAGCGCAGCAGC----- 565

QY 156 rValAlaLysSerProAlaLysThrSer-----ThrAlaLeuAsnAlaValValas 173
Db 566 -----AAGGCGCAGCGCGCCAGCACTGCTGGACCGAGCGATCTC-GCGGTGTCCAA 617

QY 173 nLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
Db 618 CTGTGCGCAGTTGATCGCGGTGAACGTCGTCAACAGCAACGCGCAGCTACAGTGTTCAT 677

QY 193 uSerTrpLysGlyAsnGlyThrProProlleserThrSerAsnTyrSerTyrAlaGlnPr 213
Db 678 GAGC-----AATGGCGCAGCGCTGCTGTGCTCCACCAAC--AGCTAC-----AA 719

QY 213 oMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAl 233
Db 720 CTGGGCGCGCGCTTCCACCGCGCAGCAGCAGTGAAGTCTGCTGCGAGTATCTCGGCA 779

QY 233 aProSerSerThrGlySerLysThrPheGluTyrLeu----- 245
Db 780 GCGCGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 839

QY 245 ----- 245

Db 840 CAGCTCGCGCGCTGCTTGTGCTTCCGCGCAGCAGCAGCTCGATCGGCGGAGCGCAACT 899

QY 246 -----ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly----- 259
Db 900 CGGCGCGATCGCGGAGCTTCTCGCGCAGCTCAACGCGCAGACGCGCTCGGTATCAC 959

QY 259 ----- 259

Db 960 GCTCGCGCGCGCAAGCGCGCGCTGTTCTCGTGGCGCGCGCGCGCGCGCTTACGCGAA 1019

QY 260 -----ThrAspSerAlaGlyLeuLeuMetSer----- 268
Db 1020 CACACAGAACACCGCGCAACGCTGCTGAACTGCTGCTGCGCGGAGCGCGCGCGCGCGAC 1079

QY 269 -----GlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThr-----Al 284
Db 1080 CACCGCGCGCTACAGCTGGCTACACGCGCAGCAGCTACACGCTGACCGACATTCGAC 1139

QY 284 aPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVa 304
Db 1140 GGGCAGCGTGGTGGTTCGCACT-----AACTGAGCGCGCG-----AT 1181

QY 304 lAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyAlaProAlaSerAlaAla 324
Db 1182 CAACGCGCTG-----AATTTCTGACCGCGCG-----ACGAT 1214

QY 324 uAspPheGly-IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344
Db 1215 GAACCGCGCGGATTCGTTACAGTCTGAGCGCGAGC-CGCGCGCGCGCTGAACAGTTTCGGA 1273

QY 344 lAlleGlyThrAspIleGlyLysLeuProSerMetMetProIle-----G 359
Db 1274 CGCGCCACCGCAGCAGTCCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1333

QY 359 lNThrSerSerGlyAsnSerThrAlaArgAsnGlySer----- 371
Db 1334 CGAGCAACCG 1393

QY 372 --SerSerThrArgArgTyrSerGlnAsp-----GlyTyrProGlnGlyA 386
Db 1394 CGAAGTCGACCGGT 1453

QY 396 spLeuAlaAspVal----- 390
Db 1454 CGACCGGTGACCGTGGCG 1513

```
Qy 391 -----ThrIleThrSerGluGlyLysLeuGlnGlyL 401
Db 1514 TGTGCGGTATTCGTGGGACCGCGGCACGCTCAGATCAACAACGCGCGCGGC 1573
Qy 401 ySyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla-----ArgPheT 419
Db 1574 AGATGAACAATGTGTGGTGACG---ATCAGCAACGACCGGCTGCGCGGCAAGTTCA 1630
Qy 419 hr-----SerGluAspGlyLeuArgArgGluGlyAsnAsn-----HisT 432
Db 1631 CCATGCGCGCGAACAACCGCGCGGACCAACGCGGTGCGAACGGCTGGCATTTGCGAATT 1690
Qy 432 yrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly----- 447
Db 1691 TGTCCACCGCGAAGCGATGTCGGGCGGAGCGGTCAAGCTGACGCGCGGTATGCGAATC 1750
Qy 448 --ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlnThrSerAsn----- 463
Db 1751 ACGTCAACAATATCGGCAAC---CAGACCAACAGATCCAGACCTCGAGCGCGCGCAGA 1807
Qy 464 -----ValAspMetSerA 468
Db 1808 GTTCGTGTTGACGACATCAACCGCGCGAGTCGTTTCGGGCGTGAACATCAACG 1867
Qy 468 rGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValT 488
Db 1868 AAGAAGCAGCGCAACCTGCTTCACTATCAGCAACTGTATCAGCGGAACAGCAAGGTCTATCC 1927
Qy 488 hrThrAlaAspThrMetLeuGlnLysAlaLeu 498
Db 1928 AGACCGCGCAGACCTGTTCCAGACGATCTC 1959

RESULT 22
US-10-193-764-64
; Sequence 64, Application US/10193764
; Publication No. US2003013943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR FILING DATE: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-193-764-64

Alignment Scores:
Pred. No.: 1,786-07 Length: 3285
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 15 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-193-764-64 (1-3285)
Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 1807 AATATTTTCAGGTTTCAATAAGACGAGATACACTAAAGAT-----GGTAGTGATTTA 1860
Qy 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 1861 ACTATTGTTACACCAATAGTGTGATGTTACTAAT----- 1896
```

```
Qy 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 1897 ---GCCAAAAAGTAACCTTT-----AACCAAGTTTAAAGAT 1929
Qy 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
Db 1930 TCAAAAATCTCTGTGACGGT---CACAAGGTGACACTACACAGCAAGTGGAA----- 1980
Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 1981 ACATCCGCTAGTAATAACAACACTGAAGATGACAGTACATAAT---GCCGGCTTAAT 2037
Qy 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 2038 ATCGATGCAAAAATGTAAACAGTAAACAACAATAATATTCTCTCAAAAGCAGTACAGCATC 2097
Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 2098 -----TTCGCCACAGTGGAGAAATACCCTCAAAACAGGTAGT-----ATC 2190
Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 2143 AACGCAACCACTGGTAACTGGAGATAACCGCTCAAAACAGGTAGT-----ATC 2190
Qy 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProIle 203
Db 2191 CTAGTGGAAATGTAGTCTCAGCTCTGGCTCTGTAACTACTGCAACCGCGGCTCTT 2250
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 2251 GGTGTAAAGCAATATTTCG-----GGCAACACC 2277
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 2278 GTTACTGTTACTGCAAAAT---AGCGGTGCTATACCACTTTGGCAGGCTCTACAAATAA 2334
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 2335 GGAACCGAGAGTGTAAACCACTTCAAGTCAATCAGCGCATATCGGGGTACGATTCTCGT 2394
Qy 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 2395 GGCACAGTAGAGTTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAATTCAAAATTTAA 2454
Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
Db 2455 GCAACAACAGCGCGGCTAAACGTAACAAGTGCACAGTCAATTTGGTGGTACGATTTC 2514
Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
Db 2515 GGTAAATACGTTAAAT-----GTTACGGCAACCGCTCGC 2547
Qy 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
Db 2548 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTAAATGCGACAGNA 2589
Qy 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 2590 GGAGCTGCAACCTTAACCTACATCATCGGCAAAATTA-----ACT 2628
Qy 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgTyrSerGlnAsp 380
Db 2629 ACCGAAGCTAGTTTCACATCTTACGCCAAGGCTAGGTAATCTTTTCAGCTCAGAT 2688
Qy 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 2689 GGTAGCGTTGACGAGGATTAATGCGCCCAATGTGACACTAAATACTACAGGCACTTTA 2748
Qy 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 2749 ACTACCGTGAAGGTTTCAACCAATTAAATGCAACCGGCTACCTTGTATTATTAACCAAAA 2808
Qy 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
```

Db 2809 GACGCTGAGCTAAATGGCGCAGCATTTGGGTAAACACACACAGTGGTAAATGCAACCAACGCA 2868
Qy : : : : :
Db 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Qy : : : : :
Db 2869 AATGGCTCGGCGAGGTAATCGGACACCTCAAGC-----AGAGTCAACATCACT 2919
Qy : : : : :
Db 447 GlyThrSerAsnTyrGlyLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Qy : : : : :
Db 2920 CGG-----GATTTAATCACATAAATGGATTA-----AATATC---ATT 2955
Qy : : : : :
Db 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
Qy : : : : :
Db 2956 TCAAAAACGGTATTAACACCGTACTGTTA---AAAGCGGTAAATTAATGATGCAATATAC 3012
Qy : : : : :
Db 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Qy : : : : :
Db 3013 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTAAGCAACGCG 3060

RESULT 23
US-10-092-880-1
; Sequence 1, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO.1
; LENGTH: 5116
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-092-880-1

Alignment Scores:
Pred. No.: 3,36e-07 Length: 5116
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 13 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-092-880-1 (1-5116)

Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 3480 AATATTCAGGTTTCATAAAGCAGAGATTACAGCTAAGAT-----GGTAGTGATTA 3533
Qy 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 3534 ACTATTGGTAACACCAATAGTGTGCTGATTAAT----- 3569
Qy 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 3570 ---GCCAAAAGTAACCTTT-----AACCAAGGTTAAAGAT 3602
Qy 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
Db 3603 TCAAAAATCTCTGTGACGGT---CACAAAGTGACACTACACAGCAAGTGGAA----- 3653
Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 3654 ACATCCGGTAGTAATACAACTGAAGATGACAGTGAACAATAAT---GCCGCTTAAC 3710

Qy 129 LeuMetGlySerArgGlyLeuSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 3711 ATCGATGCAAAAAATGTAACAGTAACCAATATTAATCTCTCAAAAGCAGTGCAGCATC 3770
Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 3771 -----TCTCGCAGCAAGTGGAGAAATACCACTAAACAGGTACCAACCAT 3815
Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 3816 AACGCAACCACTGGTAACCTGGAGATAACCGCTCAACAGGTAGT-----ATC 3863
Qy 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProIle 203
Db 3864 CTAGGTGGAATTGAGTCTCGCTCTGTAAACACTTACTGCAACCGAGGGCGCTCTT 3923
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 3924 GCTGTAAAGCAATATTCG-----GGCAACACC 3950
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 3951 GTTACTGTTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAAATTA 4007
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 4008 GGAACCGAGAGTGTAAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGT 4067
Qy 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 4068 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAAATTTAAAA 4127
Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
Db 4128 GCAACACACGCGAGGCTAACGTAACAGTGCACAGGTACAAATTTGGTGTACGATTTCC 4187
Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln 320
Db 4188 GGTAAATACGTTAAAT-----GTTACGGCAACCGCTGGC 4220
Qy 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
Db 4221 GATTTAAACAGTTGGGAATGGCGCAGAA-----ATTAATGCCACAGAA 4262
Qy 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 4263 GGAGCTGCACCTTAACCTACATCATCGGCAAAATTA-----ACT 4301
Qy 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
Db 4302 ACCGAAGCTAGTTTACACATTAATCTACTACCAAGGTCAGGTAATCTTTTCAGCTCAGAT 4361
Qy 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 4362 GGTAGCTTGCAGGAAGTATTAAATGCCGCCAATGTGACACTAATACTACAGGCATTTA 4421
Qy 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4422 ACTACCGTGAAGGTTCAACATTAATGCAACCGCGTACCTTTGGTTATTAAACGCAAA 4481
Qy 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
Db 4482 GAGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4541
Qy 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4542 AATGCTCCGGCAGCGTAATCGCAACACCTCAAGC-----AGAGTGAACATCACT 4592
Qy 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 4593 GGG-----GATTTAATCACATAAATGATTA-----AATATC---ATT 4628


```

; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE: HAEMOPHILUS
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9171
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-092-880-5

Alignment Scores:
Pred. No.: 7,75e-07 Length: 9171
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 13 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-092-880-5 (1-9171)
QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
DB 3430 AATATTTCAGTTTCATTAAGCAGAGATTACAGCTAAGAT-----GGTAGTGATTTA 3483
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
DB 3484 ACTATTGGTAACACCAATAGTGTGTGCTACTAAT----- 3519
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
DB 3520 ---GCCAAAAGTAACCTTT-----AACCAAGGTTAAAGAT 3552
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108
DB 3553 TCAAAAATCTCTGTGACGGT---CACAGGTGACACTACACAGCAAGTGAA----- 3603
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
DB 3604 ACATCCGGTAGTAATAACACACTGAAGATGACAGTGCACATAAT---GCCGGCTTAAC 3660
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlnThrLeuGluProIle 148
DB 3661 ATCGATCAAAAATGTAAACAGTAACACATAATTAATCTTCTCAAAAGCAGTGACATC 3720
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
DB 3721 -----TCTGGCACAAGTGGAGAAATTAACCACTAAACACAGGTACACCAANT 3765
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
DB 3766 AACCAACCACTGGTAACGTTGAGATTAACCGCTCAACACAGGTAGT-----ATC 3813
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProPheIle 203
DB 3814 CTAGGTGGAATTGAGTCCAGCTCTGGCTCTGTAAACACTTACTCAACCGAGGCGCTCT 3873
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
DB 3874 GCTGTAGCAATATTTCG-----GGCAACACC 3900
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
DB 3901 GTTACTGTTACTGCAAT---AGCGGTGCATTAAACCACTTTGGCAGGCTCTACAAATTA 3957

```

```

QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
DB 3958 GGAACCGAGAGTGAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGGT 4017
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGlyLeuLys 280
DB 4018 GGCACAGTAGAGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCCAATTCAAAATTA 4077
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAlaSerAlaTrpGlnPro 300
DB 4078 GCAACCAACGAGCGAGCTAAACGTAACAAGTGCACAGGTACAAATGGTGGTACGATTTCC 4137
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
DB 4138 GGTAAATACGTTAAAT-----GTTACGCGCAACGCTGCG 4170
QY 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
DB 4171 GATTTAACAGATTGGGAATGGCGCAGAA-----ATTAATCGCAGACAA 4212
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
DB 4213 GGAGCTGCAACCTTAACCTACATCATCGCGCAATTA-----ACT 4251
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
DB 4252 ACCGAAGCTAGTTACACATTAATCTTACAGCAAGGTCAGTAAATCTTTTCAGCTCAGGAT 4311
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
DB 4312 GGTACGGTTGCGGAAGTATTAAATGCCCAATGTGACACTAAATACACTACAGGCACTTTA 4371
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
DB 4372 ACTACCGTGAAGGGTTCAAACATTAATCAACACGCGGTACCTTGGTTATTAAACGCAAA 4431
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
DB 4432 GACGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGTGGTAAATGGCAACCAACGCA 4491
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
DB 4492 AATGCTCCGCGCAGCGTAAATCGCGCAACCTCAAGC-----AGAGTGAACATCACT 4542
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
DB 4543 GGG-----GATTTAATCAATAAATGATTA-----AATATC---ATT 4578
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
DB 4579 TCMAAAACCGGTATTAACACCGCTACTGTTA---AAAGCGTTAAATGATGTGAATATC 4635
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
DB 4636 ATTCACCGGCTATATGCAAGCGTAGATCAAGTAAATGAAAGCGAAACGC 4683

RESULT 26
US-10-282-122A-41760
; Sequence 41760, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

```

; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 41760
 ; LENGTH: 11118
 ; TYPE: DNA
 ; ORGANISM: Yersinia pestis
 ; US-10-282-122A-41760

Alignment Scores:
 Pred. No.: 1,3e-06 Length: 11118
 Score: 172.50 Matches: 119
 Percent Similarity: 32.30% Conservative: 58
 Best Local Similarity: 21.72% Mismatches: 180
 Query Match: 6.67% Indels: 191
 DB: 16 Gaps: 24
 US-10-009-823A-1 (1-502) x US-10-282-122A-41760 (1-11118)
 QY 3 GlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22
 DB 7075 GGGGAATTGATCTCTGGGGGGAATGGGCTTTCGGCCAGACGTCATTACTGTAATATCGCC 7134
 QY 23 SerAsnAsnIleAlaAlaAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGln 42
 DB 7135 AGTGGG-----GCCAGTGGCAATATTAAATGGTATCGCCAG----- 7170
 QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAla 62
 DB 7171 -----ACGGTGGGGGGCGGTGACCAATACCGGTACGGTAAACGTTG 7209
 QY 63 GlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluPro 82
 DB 7210 GGCACCGTGGGAGTTAACAGTCTGACACCTTGATCAATACCGGAATGATTATG 7269
 QY 83 GlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGlu 102
 DB 7270 ACCGATCGCATCTGTAATCTGGAGATGGGGGCTTCTAGCATTAGCGCGCTTA--- 7326
 QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
 DB 7327 -----ACGGCAACCGGTATCTGGAAT 7347
 QY 123 AspProSerGly---PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLys 141
 DB 7348 ATCAAGGGTGGGATTTTACCATC-----AGCATCGATAACAAAT----- 7386
 QY 142 LysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerPro 161

Db 7386 ----- 7386
 QY 162 AlalysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
 Db 7387 -----GGTCTGGCGGGCAACCAATATTTCCGATGGTGCATCATGCTACT 7431
 QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
 Db 7432 -----CTTGGTAAATGGGGGACC 7449
 QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGly 221
 Db 7450 ATTATCGGAACCGGTAAT----- 7467
 QY 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241
 Db 7468 -----TTGGGAGCAGCGCTT 7482
 QY 242 PheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSer----- 258
 Db 7483 ATTGATGTCTGGGGGATCTAAACCTGTCTCGCGGATAATTCTCTGGCTAACGTCATCAGT 7542
 QY 259 -----GlyThr-----AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
 Db 7543 GGTGACGGGACGATTAAATACCAACAGACAGTACGCTGACGGGAATAGCAGCTTTAGT 7602
 QY 275 -----SerAsnGlyGluLeu-----LysAsnMetThr 283
 Db 7603 GGTGCACATCAGATTGGTACCAATGGCGAATGACCGTGGGTGAGCCAGTAATCTGGGG 7662
 QY 284 AlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeu 303
 Db 7663 GCC-----AGCAGCGCCACCGTTAATCTGGGCACCCCTTACTCTCATCTGATC 7710
 QY 304 ValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThr 323
 Db 7711 TTGATGGCGTTCGGAGAGTATTGCCAAC---GTTCTGACGGGTGTGGCGGTTCAACA 7767
 QY 324 LeuAsp----- 325
 Db 7768 GTCGATATTATCGCGGAGCAGATACCGACTGACGCCCAATAACAGCGGCTTCTCTCGGC 7827
 QY 326 -----PheGlyIleLysSerGlnGlnAsnMetTrpAla 336
 Db 7828 CAATATGCTTGGCGGGTAAACAGCAACTGACGGTGGCTCAACGAAACATCTG----- 7881
 QY 337 GlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
 Db 7882 GGGGCGTTCATCCAGCGTGGCGTGGCAGGAGC-----GGCGATACCCCTCTCGCTGAGC 7935
 QY 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArg 376
 Db 7936 GCGTTTAAACGGCACCTTTGGTAAACAGCGTCAACCGGACGGGTGACTGCAAGTGAAC--- 7992
 QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGly 396
 Db 7993 -----GATGATGCCGAAGTCAACCTGACCCAGCAGCAAC 8025
 QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu--- 415
 Db 8026 GGGGTA-----AGCAACGCGGTAATACTGATATCGCTGATCGCAGCGCTAAAT 8073
 QY 416 -----AlaArgPheThr-----SerGluAspGlyLeuArgArgGluGly 428
 Db 8074 CTGGACGATATTGCTCTCTTTAATCATCGCTTACCGGTAAACGGCTTGTCTGAATGTGCG 8133
 QY 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly---LeuProGly 447
 Db 8134 AAAACAGTATGCGACGCGCTTTGACTTGTGCTACCGTGGGGGGCGCTTCAACCGC 8193
 QY 448 ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer 467

QY 112 AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131
 Db 4102 AACTGAGCTTCACCGCGCTCGCAACTGCGCAACGCGACCGTGTCAACGTGACCGCG 4161
 QY 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp 151
 Db 4162 AGCAGCGCGCGCAATACCAAGCTTCGCGTACCAGC-----ACGGTGGAT 4209
 QY 152 PheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 171
 Db 4210 TCCTCGCTGCGTCCGATCCGCGAGGTGGATTCGAGCAACGGTTCGGTGATCAGCGCAC 4269
 QY 172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu 191
 Db 4270 CGCAGCGCGCGCAACCATCATCATCAGCGGCAACGGCAACCGGATTCGGCCAGGTC 4329
 QY 192 LeuGluSerTrpLysGlyAsnGly-----ThrProIleSerThrSerAsnTyrSer 209
 Db 4330 ACCGCGCAGCGCGCAACTGCTTCCTTCACTCCAGGATCCGCTGCGCGATGCGCAC 4389
 QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229
 Db 4390 GTGTCACGTGTGGCGCGCCAGCCAGCAAGCATGTGACAGTCCGCGCGGGTGATCAT 4449
 QY 230 PheAspGly---AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
 Db 4450 GTGATGCGTGGCGCGCGCGCGCGC-----GTGATC 4482
 QY 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet--- 267
 Db 4483 GATCCGAGC---AACGCGACCGAGATAGCGGTACCGCGAGCGCGCGCGCGGATGATC 4539
 QY 268 -----SerGlyThrMet 271
 Db 4540 CTCACGATGGCGCGCAACCGATCGCGCGAGCGCACCGCGCGCGCAACTGG 4599
 QY 272 ThrPheSer-----SerAsnGlyGluLeuLysAsnMetThrAlaPheThr 286
 Db 4600 ACCTTACCCCGGACGACCCCGCTGGCGCAACCGCGTGTATCAACCGCGTGGCGCGAG 4659
 QY 287 ProThrGly-----SerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
 Db 4660 CCGCGCGCAATACCGAGCGTGGCGCGAGCGTCACTGATGCCATCGCC---CCGCGG 4716
 QY 301 AlaProLeuValAsn-----GlyLeuProGlnPheSerAla 312
 Db 4717 GCGCGGTGATCAATCCGAGCAATGGATGCTCATCAGCGGTACGCGGGAAGCGCGGCGC 4776
 QY 313 AsnPheVal-----GlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327
 Db 4777 ACGTGATCTCTACCGACGCAACCGCAACCGATCGCGCGAGTCAACCGCGAC----- 4830
 QY 328 IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr 347
 Db 4831 -----GCGACGCGCAACTGGAGCTTCAGCGCGCGCACCGCTGGCGCAACGCGTGG 4881
 QY 348 AspIleGlyLysLeu----- 352
 Db 4882 GTGATCAATGCTGCGCGCGCGAGCGCGCGCGCAACACAGCATGCCACCGCGCAC 4941
 QY 353 -----ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla 367
 Db 4942 GTGACTGCTGCGCGCGCGAGCGCGCGGTGATCGATCCGAGCAACGGTACGGTATCGCC 5001
 QY 368 ArgAsnGlySerSerThrArgArgTyrSerGlnAsp-----GlyTyrProGlnGly 395
 Db 5002 GGTACCGCGCGAGGTGGTGGCGCGCGTATCTCACCAGCGCAACCGCATCGCGC 5061
 QY 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 Db 5062 CAG-----GTCACCGCGGATGGC-----AGCGGCAACTGGAGCTTCACG 5100

QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
 Db 5101 CCGCGCAGCGCGCTGTCCATGTCACGCGTGTCAATGCGTGGCCAGGACGCT----- 5154
 QY 426 ArgGluGlyAsnAsnHis-----TyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
 Db 5155 ---GCGCGCAACACCGCGCGCGCTCAGCACCGCGTGGAGCG---GGTGGCCCC- 5204
 QY 444 GlyLeuProGly 447
 Db 5205 GCGCACCGCGGT 5216

RESULT 29
 US-10-282-122A-30151
 ; Sequence 30151, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zvakind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA 034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 30151
 ; LENGTH: 7407
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-10-282-122A-30151

Alignment Scores:
 Pred. No.: 3,466-06 Length: 7407
 Score: 166.00 Matches: 113
 Percent Similarity: 25.12% Conservative: 64
 Best Local Similarity: 22.42% Mismatches: 189
 Query Match: 6.42% Indels: 139
 DB: 16 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-282-122A-30151 (1-7407)
 QY 20 GlyThrValSerAsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnVal 39

3892 GGACGGTGGTCAACGGCGTGCC----- 3915
Db
40 ValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGly--- 58
Qy
3916 -----CAGGACCTCGG--GGCAATACCGCCCGCAGGGCAGC 3951
Db
59 -----ProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal 71
Qy
3952 ACTACCGTGGAGCGGTGGCGCGCAACACGCGCTGTGTCAATCCGAGCAACGCACTG 4011
Db
72 ArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIle 91
Qy
4012 -----CTCAACGGTACCGCGGACCGCGGCGACACCGTGATCTGACCGAGCGG 4059
Db
92 GlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGly 111
Qy
4060 AAGGCAACCGCATCGCGCACACCGCGCGATGCG-----AGCGGC 4101
Db
112 AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131
Qy
4102 AACTGAGCTTCACGCCCGGCTCGCACTGCCCAACGCGCACCGGTGTCACGTGACCGCG 4161
Db
132 SerArgIleSerAsnAspProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp 151
Qy
4162 AGCGAGCGCGCGGCAATACCGACGCTTCCTCCGCTACCG-----ACGGTGGAT 4209
Db
152 PheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 171
Qy
4210 TCCTCGCTCGCGTGCATCCCGCAGGTGGATCCGAGCAACGCTTCGCGTGCATCGCGG 4269
Db
172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu 191
Qy
4270 CGGAGCGCGGCAACCATCATCATCCGATGGCAACGGCAACCGATGGCCAGGTC 4329
Db
192 LeuGluSerTrpLysGlyAsnGly-----ThrProIleSerThrSerAsnTyrSer 209
Qy
4330 ACCGCGCAGCGCAGCGCACTGGCTCTTCACTCCAGCGCATCCCGTCCGCGATGGCAGC 4389
Db
210 TyrAlaGlnProMetArgValTyrAspGlnGlnLysSerHisAspIleThrValTyr 229
Qy
4390 GTGGTCAACGTGGTGGCGCGCAGCCCAAGCATGTCACAGCTGCGCGCGGTGATCACT 4449
Db
230 PheAspGly---AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
Qy
4450 GTGGATGGCGTGGCGCGCGCGCGCG-----GTGATC 4482
Db
249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet--- 267
Qy
4483 GATCCGAGC---AACGCGACCGAGATAAGCGGTACCGCGAGCGCGCGCGGTGATC 4539
Db
268 -----SerGlyThrMet 271
Qy
4540 CTCACCGATGGCGGCAACCGCATCGCGCAGCGCACCGCGCAGCGCACTGG 4599
Db
272 ThrPheSer-----SerAsnGlyLeuLysAsnMetThrAlaPheThr 286
Qy
4600 ACGTTCACCGCGAGCACCGCGTGGCGCAACCGCATCGCGCGGCGCGCGCGGCG 4659
Db
287 ProThrGly-----SerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Qy
4660 CCGCGCGGCAATACAGCGGTCCCGCGCAGCGTCACTCGATCGATCGCGGCGGCG 4716
Db
301 AlaProLeuValAsn-----GlyLeuProGlnPheSerAla 312
Qy
4717 GCGCGGTGATCAATCCGAGCAATGGAGTGGTCACTACGCGGTACCGCGGAAGCGCGG 4776
Db
313 AsnPheVal-----GlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327
Qy
4777 ACGGTGATCTCCACCGAGCGCAACCGGATCCGCGAGGTACCGCGCGGCGGCG 4830
Db
328 IleLysSerGlnGlnAsnMetTrpAlaGlyValProAlaSerAlaAlaIleGlyThr 347
Qy

4831 -----GGCAGCGGCAACTGGAGCTTCAAGCCCGGCGACGCGCTGGCCAAAGGCTCG 4881
Db
348 AspIleGlyLysLeu----- 352
Qy
4882 GTGATCAATCGCTGGCCAGGACCGCGCGCAACACAGCAGTCCACACGCGCCACC 4941
Db
353 -----ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla 367
Qy
4942 GTGACTCGCTGGCGCGCAGCAGCCCGGTGATCGATCCGAGCAACGCTAGCTGATCGCC 5001
Db
368 ArgAsnGlySerSerThrArgArgTyrSerGlnAsp-----GlyTyrProGlnGly 385
Qy
5002 GGTACCGCGGAGCTGGTCCACGCGTGTCTCACCGGCAACGCGCAACCGCATCGCG 5061
Db
386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
Qy
5062 CAG-----GTCACCGCGCATGCG-----AGCGGCAACTGGAGCTTCACG 5100
Db
406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
Qy
5101 CCGGCGACCGCTCTCCATGGCAGCGTGTCAATGCGGTGGCGCGCGCGCGCT----- 5154
Db
426 ArgGluGlyAsnAsnHis-----TyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Qy
5155 ---GCCGCGCAACACCGCGCGCGCGTGCAGCACCGGTGGACGC---GGTGGCGCC 5204
Db
444 GlyLeuProGly 447
Qy
5205 GCCCACCACCGGT 5216
Db
RESULT 30
US-10-335-977-77
; Sequence 77, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...372
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-335-977-77

Alignment Scores:
Pred. No.: 5.39e-08 Length: 372
Score: 165.50 Matches: 39
Percent Similarity: 51.64% Conservatives: 24
Best Local Similarity: 31.97% Mismatches: 48
Query Match: 6.40% Indels: 11
DB: 16 Gaps: 3

US-10-009-823A-1 (1-502) x US-10-335-977-77 (1-372)

QY 392 IleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyr 411
Db 7 ATTGAGAAATGCGGTATTTCTTAGCCTTTAGTAACGGGTGTAGAGCGGTCGCT 66
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnHis 431
Db 67 CGCATCGGTATTWAGCTTTCACTACGATCAAGGCTTAAGGAAATCGCGTAACCTC 126
QY 432 Tyr-----SerAlaThrLeuAsp-----SerGlyGlyProGlu 442
Db 127 TATGAATGCAAGAGCCATTAATGCGGAAACAGACCCCTAAGSGTAAACCCAT 186
QY 443 PheGlyLeu-----ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
Db 187 TTAGGGTGGGACGAGAGGCGCAAGCTCAAGCTTTGGGAAATCAGGCACAAATATTAGAA 246
QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
Db 247 ACAGAGCAACGTGAATGCGGGAACGCCCTAACCAATCTCATTTAATGCAAGAGCTAT 306
QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu 500
Db 307 TCTATGACGCTAGAGCTTTGGCGGCGGATGACATGATCAAGAGCAATAGCTTG 366
QY 501 LysArg 502
Db 367 AAAAAA 372

RESULT 31

US-10-114-170-39

Sequence 39, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESS: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 25165
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-114-170-39

Alignment Scores:
Pred. No.: 2.53e-05 Length: 25165
Score: 165.00 Matches: 145
Percent Similarity: 33.70% Conservatives: 71
Best Local Similarity: 22.62% Mismatches: 189
Query Match: 6.38% Indels: 236
DB: 14 Gaps: 35

US-10-009-823A-1 (1-502) x US-10-114-170-39 (1-25165)

QY 8 GlyAlaThrGlyMetLysThrHis-----SerThrGlyLeuGlyThrValSer 23
Db 13018 GGAACAGCGGCGAGCGGTACCCATCAGGTGAGCGGTCAATACCGGCTCCGACCATACC 13077
QY 24 -----AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGln 37
Db 13078 TTAAAGCCGATCAGCGGCGATAACATCTGAAACGCCGATGAAAGGC----- 13125
QY 38 GlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIle-----GlySerThrGly 55
Db 13126 -----CAGCGGTTGACCATCAGCGGCGGCGAGTACGGGG 13158
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 13159 -----CTGGCGAGCGGCGCGAGGTC-----ACCGTCAGG 13188
QY 76 ThrGlnGly-----AlaPheGluProGlyAsn----- 84
Db 13189 CTCACGCTCACAACTACAGCGCCACCCAGCGCATCGGCAACTGGACCTTAACCGTG 13248
QY 85 SerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLys 104
Db 13249 CCGGTGAGCGATCTGCGGCATTAGGTGAGGCAACTATACGTCAGCGCGAGCCGCCACC 13308
QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124
Db 13309 -----AGTGCAGCAGCGCAACCCGCGCAGCGCGGCGCAATTACTGCTCGAC--- 13356
QY 125 SerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThr 144
Db 13357 -----AGCGGCTGCGCGAGGTCACCATCAACACC 13386
QY 145 LeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThr 164
Db 13387 GTGGCA-----GGCAGCATATTATCAACGCCGCCGCGGCGGGGCC 13428
QY 165 SerThrAlaLeuAsnAlaValValAsn-----LeuGlyAspSerThrAspLysThr 181
Db 13429 GATCAAAACCATCAGCGGCGGTGTGACTCGCGCGCGCTGGCGATACGTCACCGTAGCG 13488

QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
Db 13489 CTGGGC---GGGAACACTTACACCGCTACGGTACAG----- 13521
QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnIn--- 220
Db 13522 -----AGCAATTAGCTTGGAGCTGACGGTTCGACAGCGGATCTCCAGCG 13569
QY 221 ---GlyAsnSerHis---AspIleThrValTyrPheAspGlyAlaProSerSerThrGly 238
Db 13570 TTGGGCAATGGTGAATTGACCATTTACCGCTCGGTCCGTCACCAACGCTAATGGCAACACCGGG 13629
QY 239 SerLysThrPheGluTyrLeuVal----- 246
Db 13630 AGCGCACGCGGATATCACCATGTATGCCAACCTGCCGGGGCTGGCGGTAGATACCGTG 13689
QY 247 -----AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThr---Asp 261
Db 13690 GCGGCGATGATATCGTCAACAGCATCGAGCAGCGGAGGCGGTGTGATCACCGCGCGC 13749
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281
Db 13750 AGTACGCGCCTGAATGACAGGTGCTGTGCTGACGGTTACCATCAACAGT----- 13797
QY 282 MetThrAlaPheThrProThr-----GlySerAlaThr 292
Db 13798 ---GTGGCGTATTCGCGCACCGTGCAGCGGAGCGAGCTGGAGCTGGCATTCGCGCG 13854
QY 293 LysAspLeuAlaTrrpGlnProAlaProLeu-----ValAsnGlyLeuProGln 309
Db 13855 GCAACAGTCACGCGCTCGCTCGCGGCGGCTTAAACCGTGGAGGTAGACGG---CAA 13908
QY 310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe----- 326
Db 13909 AGCAGCGCCATTAACCCAGTCAGCGTCAGCATCCGTTCCGTCGATTTAACGGCGGTG 13968
QY 326 ----- 326
Db 13969 GCAATCAGCATCAACACCGTTTCCAGCGAGCAGCGTGATTAAACGCGCAGAAAAAGGCACC 14028
QY 327 -----GlyTyrLeuSerGlnGlnAsnMet----- 334
Db 14029 AATCTGACTCTTTCGGCAGTACAGCGGATTTAGAGCGGCAAAACCGTCACCGTCAC 14088
QY 335 TrpAlaGlyAla-----ProAlaSerAlaAlaIleGlyThr----- 347
Db 14089 TTTGGCGTAAACCTACACTGCAAGCGTCGCGCACAACGGAGCTGGAGTGTAAACGTT 14148
QY 348 -----AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 14149 CCGCGGCGAGATCTGGCAACTCTGCCAGAGGCGCGCGGCAATGTGACGCGCAGCGTTAGC 14208
QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp----- 380
Db 14209 AGCGGAGCGGTACAGTGCCTCGCGACCCATCGGTATAGCTTGACGCCAGCGCGCG 14268
QY 381 -----GlyTyr 382
Db 14269 ACGTCTACCATTAACACCATCGCGCAGCATATCTTAAACCGCAGAACCGGAAGC 14328
QY 383 Pro-----GlnGlyAspLeuValAspValThrIle 392
Db 14329 CCGCTCACCATCAGCGCACACGACCGCGGAACCGGCGAGAGCGGTGACCGTCAACCTT 14388
QY 393 -----ThrSerGluGlyLysLeuGln-----GlyLysTyrSerAsnSerGlnVal 407
Db 14389 AACGGCGCAACCTACACCGGCACTGTGCGAGCGGAGCGGTAGCTGGAGCGGTACG--- 14442
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 14443 -----GTTCCC-----ACTTCAGCCCTGGCGCGGTCAACGCA 14475

QY 428 GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447
Db 14476 AGCAATTATACCGTCCAGCCGACGGTCAATGAC-----AAAGCGGCAACCCCGGC 14526
QY 448 ThrSerAsnTyrGly-----LysLeuSerValAsnGlnLeu 459
Db 14527 AGCGCGACCATTAATCTGGCGGTAGACACCGCGCGCGGTCTCCATTAAACCGGTG 14586
QY 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArg--- 478
Db 14587 GCGGCGGATGATCATCATCAACGATCCGACATGCGGAGGCGTGTGATCTCCGCGCAC 14646
QY 479 -----GlyPheGlnMetAsnSerLysSerValThr 488
Db 14647 AGTAGCGCGGGAAGCGGCGCATGTGTGAGCGTGTGCTCAACGCGCAACCTACACC 14706
QY 489 Thr 489
Db 14707 ACC 14709
RESULT 32
US-10-335-977-57
; Sequence 57, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1029
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-335-977-57

Alignment Scores:

QY 129 LeuMetGlySerArgIleSerAsnProAsnLeuLysLysGluThrLeu----- 145
DB 1688 CTGACTTCTCAAGTAACCTTGAATCTCCCAACGAGACAGAGTACATCAAGCAACGCA 1747
QY 146 -----GluProIleGlnLeuAspPheAsnAspProThr----- 156
DB 1748 AACGGAACCTCAAGGAATGCTTGGCTCTCTGTCGAACTCAGAACATACAGCAACG 1807
QY 157 -----ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn--- 173
DB 1808 TTAGTGGGTTACTTCTTCCGCTACCACTCTTACGCTCAAAACGCGGTTTAAATTCG 1867
QY 174 -----LeuGlyAspSerThrAspLysThr--- 181
DB 1868 ATCAGATTACTCCGCTCTCTTCCGTCAGTAGAGGAAGTCTGAACCTTACTGCA 1927
QY 182 -----GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGly 197
DB 1928 ACCGGAACCTTATCTGACGGAAGTAACCAAGATTAACTACCTCCGTCGCTTGACGAGT 1987
QY 198 AsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyr 217
DB 1988 ACGGATCTTCCATCGTTCTCGTAGACACGCTCAGTAGACAGGCGGACAGCAGGT 2047
QY 218 AspGlnGlnGlyAsnSerHis----- 224
DB 2048 GTTGCAACAGGTAACACTCAGATCAGTGCACATAGGCGGAACCTTCTTCTGTATCAAT 2107
QY 225 -----AspIleThrValTyrPheAspGlyAlaProSer 235
DB 2108 TTTACGTAAGTGCAGCGGTTTATAGATTCAATTCAGTAATCTCGAAGATTCTCCGATT 2167
QY 236 SerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySer 255
DB 2168 GCAAAAGGAACCTTCTACA-----AGAGCAATCGCAGCGGTGTTTTTTCAGACGGAAGC 2221
QY 256 AlaAlaSerGlyThrAspSerAla----- 263
DB 2222 AATTGATATATAGTATGATCAAGTTATTGGATAGTTTCAAAACAAACGTGATCCAGCTA 2281
QY 264 GlyLeuLeuMetSerGly-----ThrMetThrPheSerSerAsnGlyGluLeu 279
DB 2282 GGAGTTTGAAGAACCGGCTCTAAAGAACTGATGATTCCTCCGCAATGCA----- 2335
QY 280 LysAsnMetThrAlaPheThrProThrGlySerAlaThr----- 292
DB 2336 ---AACAGTACCTGGAACCTCAGGATCACTGCAACGTTAGAGGTGTGAGCGGATAC 2392
QY 293 LysAspLeuAsnAlaTyrGlnProAla----- 301
DB 2393 GCCGATCTTACAGTAATCGCTCCAGTTTAAACAGCATTCAATCGATCTCACATCCG 2452
QY 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
DB 2453 AGCGTTGCAACGCTCTGACTCAA-----AATTTTACTGCAACCGGAGTT----- 2497
QY 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet-----TrpAla 336
DB 2498 -----TACTCAGATGTAGCAATCAAGATTCAACCGATTCCGTTACTTGGCG 2545
QY 337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
DB 2546 TCTTCCAATCTGCTGTGTCACGATCAGCAACGCTTCCGGAACCAACGGAAGTACT 2605
QY 357 ProIleGlnThrSerGlyAsnSerThrAlaAsnGlnGlySerSerSerThrArgArg 376
DB 2606 ACTCTTCAAACTGATCCCAATATCAGCGGAGTCTGGCGGCCACTACTTCT----- 2659
QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlyGly 396
DB 2660 -----GATCCAAGTGTATTAAACGGTTACA----- 2683
QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416

DB 2684 -----AACGCAACCTTAAACAAGTATACAGATCGTCCACCTCT 2722
QY 417 ArgPheThrSerGluAspGlyLeuArgGlu----- 427
DB 2723 TCCTTTCAACATCGCAAAAGGATTAATCAAGACTTTTGTAGCGACCGTTATTATACAGAT 2782
QY 428 -----GlyAsnAsnHisTyrSerAla 434
DB 2783 GGTTCCTTAGAGACTGACCACTCAAGTCACTTGGAAATCTTCCATPACTTCTACCCGT 2842
QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
DB 2843 ACGATCAGCAATGCAACGGAACCTCAAGGAAGAATGGCGGTCGATCTGTTCTACA 2902
QY 455 SerValAsn-----GlnLeuGluThrSerAsnValAspMetSerArg 468
DB 2903 AATATCTCCGCTCTTTAGAGGAACGATATAGTCAGACCAACGTAACCGTTACATCT 2962
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
DB 2963 GCGGTTCTGAATTCGATCCAGTTTCTCCAGCGGACATTAGTGTAGCCAAAGGAACACC 3022
QY 489 ThrAlaAspThrMetLeu 494
DB 3023 AAGGCTTACACCGCATC 3040
RESULT 34
US-10-282-122A-22709
Sequence 22709, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22709
LENGTH: 3693

TYPE: DNA
ORGANISM: Helicobacter pylori
US-10-282-122A-22709

Alignment Scores:
Pred. No.: 5,4e-06 Length: 3693
Score: 160.00 Matches: 115
Percent Similarity: 35.93% Conservative: 74
Best Local Similarity: 21.86% Mismatches: 179
Query Match: 6.19% Indels: 158
DB: 16 Gaps: 26

US-10-009-823A-1 (1-502) x US-10-282-122A-22709 (1-3693)

```
QY 20 GlyThrValSerAsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnVal 39
Db 1624 GGGACTACCACTAATACTAAGCTAAAGCAAGCGCTCCCAACTAAAGCGATGGTGG 1683
QY 40 Val-----PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer---ThrGlySer 56
Db 1684 GTGAATAATGAAGAAGAGCTAAAGCGCAATTAGCCAAAGCAGCGGACCAACA 1743
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 1744 CAATCTCTTAACGACGCGGTGATGGAGCT-----TTAAACACCGGTGTTG--- 1788
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
Db 1789 -----CAAAATGTCAGCAAT----- 1803
QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
Db 1804 TTCCAAACAAAGCATTT----- 1818
QY 117 GlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsn 136
Db 1819 CAAAACCTTTTCAAAACCAAGAAAGTAATATCCAGCTTGGCGGATGCGATTATAAC 1878
QY 137 AsnProAsnIleLysLysGluThrLeuGluPro-----IleGln 149
Db 1879 ACTAATGGGAGTCAGTCGCAAGAGATGACACCTAACCAATAACCAAGATTTACGCATCCAA 1938
QY 150 LeuAspPheAsnAspProThrValAlaLysSer---ProAlaLysThrSerThrAlaLeu 168
Db 1939 TTGAGGCGCAATTTTACCAGCTCATCAATACCAATTAACAGCAGTGCGCTACAGCATG 1998
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 1999 AATGCTTTAATATCAAGCCCAACCAACCAACCAAGCGATCAGCAAGCAATAAT 2058
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr 208
Db 2059 AACGCATGCGCGAGTGAATGAGTGGG-----AGTAATGGTAACCTGG 2100
QY 209 SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrVal 228
Db 2101 TGCTATACGCAA-----TGTCGGATCTCAGGCT 2130
QY 229 TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
Db 2131 TATTACAGCGGTTGCAAGCGCTTACGGTATCAAAAG----- 2169
QY 249 AsnProSerGluAspGlySerAlaAspSerGlyThrAspSerAlaGlyLeuLeuMetSer 268
Db 2170 -----CAAGCCGACAACTCAAGCGGGAGCAATGTGGGAACAGCATCACCTAC 2217
QY 269 GlyThr-----MetThrPheSerSerAsnGlyGluLeuLys----- 280
Db 2218 AATGTCACCAAAATCACGCTCACTAGTAATGGTTGCTCAACCAAAATCATCAAAATCTT 2277
QY 281 -----AsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 2278 AAGAGCGTTAATGGAGGCAATGGCGGAGTGGTACAGGCAGTGGGAATGGCCAGTCAA 2337
```

```
QY 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 2338 ATCAAC----- 2343
QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 2344 -----ACAGCGCTACAGATGCTCACA----- 2364
QY 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 2365 -----GAGCGCAGCGATGGGAATTAGGACT----- 2391
QY 355 MetMetProIleGlnThrSerSerGly-----AsnSerThr 366
Db 2392 -----TATAGTAGTAGTAGTGGCAGTAATAACCGCTATATACGCCCATGCAATAGACCC 2442
QY 367 AlaArgAsnGlySerSerThrArg---ArgTyrSerGlnAspGlyTyrProGlnGly 385
Db 2443 -----AATGGAGCAATAAACAAGTGGGACAAATTGTTATGATGAACCAACAAACAA 2496
QY 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
Db 2497 AAGGCCACCCAGCAACCGCCACCAACCGCAGCAATTTACAAAAAGTCTATAATGACGCC 2556
QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer-----GluAsp 422
Db 2557 CAAAAAATA-----GCCAACATATCCCGAGCTCTCGGAACAATAAAGCGGTGAAAC 2610
QY 423 GlyLeuArgArg-----GluGlyAsnAsnHisTyrSerAlaThrLeuAspSer----- 438
Db 2611 GGTATAAACCAATCTTTGAAGCGTTAAAAATAATAGCAGCAGTCTCAGTAATTTATGT 2670
QY 439 GlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu-----SerVal 456
Db 2671 GGTAAATGTTAGTAGCGGTAGTAGTGGCAGTACTTCTCGGTGGCTTATCAACCTTTA 2730
QY 457 AsnGlnLeuThrSerAsnValAspMetSerArgGluMetValAsnMetIle----- 474
Db 2731 GGGGCAATCCCAACCAATGAGTGGAGCGATAGCAATAATTAATTAATCTGCTCACTGAA 2790
QY 475 -----IleIleGlnArgGlyPhe---GlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 2791 TTCAATTAACCGCGGGTTTATCCAAATAATGATAGTAGTGTATCTACTAGTCTTACA 2850
QY 493 MetLeuGlnLysAlaLeu 498
Db 2851 AGCGCTTTTCAAGCCATT 2868
```

RESULT 35

US-09-881-752A-149
Sequence 149, Application US/09881752A
Patent No. US20020115078A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/041002
CURRENT APPLICATION NUMBER: US/09/881,752A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/833,457
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 370
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 149
LENGTH: 3793
TYPE: DNA
ORGANISM: Helicobacter pylori

FEATURE:
 NAME/KEY: CDS
 LOCATION: (51)...(3740)
 US-09-881-752A-149

Alignment Scores:

Pred. No.: 5,618-06 Length: 3793
 Score: 160.00 Matches: 115
 Percent Similarity: 35.93% Conservative: 74
 Best Local Similarity: 21.86% Mismatches: 179
 Query Match: 6.19% Indels: 158
 DB: 26 Gaps: 26

US-10-009-823A-1 (1-502) x US-09-881-752A-149 (1-3793)

QY 20 GlyThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnVal 39
 DB 1674 GGGACTACCACTAATPACTCAAGCTAAAGCAACGCTCCCAACTAAAGCGATGGTGTG 1733
 QY 40 Val-----PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer---ThrGlySer 56
 DB 1734 GTGAATATGAGAGAGAGCTAAACGGCCCAATTAGCCCAAGCAGCGGACCAACACA 1793
 QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 DB 1794 CAATCTCTAAACAGCAGCGTGTGGAGCT-----TTAAACACCGGTGTG--- 1838
 QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysPhe 96
 DB 1839 -----CAAAATGTCAGCAAT----- 1853
 QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
 DB 1854 TTCCAACAAGCAAT----- 1868
 QY 117 GlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsn 136
 DB 1869 CAAACCGCTTTTCAAAACCAAGAAAGTAATATCAAGCTTGGCGAATGCGATTATAAC 1928
 QY 137 AsnProAsnIleLysLysGluThrLeuGluPro-----IleGln 149
 DB 1929 ACTAATGGGAGTCTAGTCGAGAGATGACACCTTAAACAATACCAAGATTACGCATCCAA 1988
 QY 150 LeuAspPheAsnAspProThrValAlaLysSer---ProAlaLysThrSerThrAlaLeu 168
 DB 1989 TTGAGGGCGAATTTTACCAGCTCATCAATACCATTAAACCAAGCAAGTGCCTACAGACATG 2048
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 DB 2049 AATGCTTTAATATCAAAACCAACCAACCAACCAACCAACCAACCAACCAACCAATAT 2108
 QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProIleSerThrSerAsnTyr 208
 DB 2109 AACCGTCGCGAGTGAATGAGTGGG-----AGTAATGTTACTCG 2150
 QY 209 SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrVal 228
 DB 2151 TGCTATCAGCAA-----TGGTCCGATTTCTAAGGCT 2180
 QY 229 TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
 DB 2181 TATTACAGCGGTTGCAACGCGCTTTAGGGTATCAACG----- 2219
 QY 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer 268
 DB 2220 -----CAAGCGACAATCAAGCGGAGCAATGGTGGGAACAGCATCCTAC 2267
 QY 269 GlyThr-----MetThrPheSerSerAsnGlyGluLeuLys----- 280
 DB 2268 AATGTCACAACATCAGCTCATAGTAATGGTTTCTCAACCAATATCATCAATCTT 2327
 QY 281 -----AsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294

DB 2328 AAGAGCGTTAATGGAGCAATGCGCGAGTGTGTACAGGCAGTGGGATGGCCACGATCAA 2387
 QY 295 LeuAsnAlaTyrPheGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
 DB 2388 ATCAAC----- 2393
 QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
 DB 2394 -----ACAGCCTACCATGCTCTACA----- 2414
 QY 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
 DB 2415 -----GAGCCAGCGATGGGAANTTAGGAGCT----- 2441
 QY 355 MetMetProIleGlnThrSerSerGly-----AsnSerThr 366
 DB 2442 -----TATAGTAGTAGTAGTGGCAATTAACGGCTATACGCCATGCAATAGCACC 2492
 QY 367 AlaArgAsnGlySerSerSerThrArg---ArgTyrSerGlnAspGlyTyrProGlnGly 385
 DB 2493 -----AATGGAGCAATAAACAAGTGGGAACAATTGTTATGAACCAACCAACAA 2546
 QY 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 DB 2547 AACGCCACCCAGCAACCGCCACCAACCGACAGCAATTTTACAAAAGTCTATATGACGCC 2606
 QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer-----GluAsp 422
 DB 2607 CAAAAATA-----GCCAACATTATCCGAGCTCTGGGAACAATAAAGGCGTTGAAC 2660
 QY 423 GlyLeuArgArg-----GluGlyAsnAsnHisTyrSerAlaThrLeuAspSer----- 438
 DB 2661 GCCTTAAACAATCTTTGAAGCGTTAAATAATATAGCAGCAGCTCTCAGTAATTTATGT 2720
 QY 439 GlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu-----SerVal 456
 DB 2721 GGTATGTGTAGTAGCGGTAGTAGTGCGCACTACTTGTCCCGTTGGCTTATCAACCTTTA 2780
 QY 457 AsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle----- 474
 DB 2781 GGGGCAATCCCAACCAATGGAGTAGGAGTAGGAGTAGGAGTAGGAGTAGGAGTAGGAG 2840
 QY 475 ---IleIleGlnArgGlyPhe---GlnMetAsnSerLysSerValThrThrAlaAspThr 492
 DB 2841 TTCATTAACCGCGCGGCTTTATCCAAAATAATGATAGTAGTAGTAGTAGTAGTAGTAG 2900
 QY 493 MetLeuGlnLysAlaLeu 498
 DB 2901 AGCGCTTTTCAAGCCATT 2918

RESULT 36

US-10-282-122A-12963

Sequence 12963, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: EUTPA 034A

CURRENT APPLICATION NUMBER: US/10/282,122A

PRIOR FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

Db 8007 CTTCTCGTGGATGCAAGAGCTACAAAGTGTCTGCTACGACATGAGCGGTGTAACGC 8066
 Qy 413 eProleulalaRgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSe 433
 Db 8067 CAAGATCGCAACG---GGATCGACGAGCGGGTGTGTGTACGACACGTCGGCACATAACAA 8123
 Qy 433 rAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro----- 446
 Db 8124 GTTGAGCTC-----GGCGAGTCAATGCCACACGCGGTGAGGTGCAACGTCGC 8177
 Qy 447 ---GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSer-----As 463
 Db 8178 CGCCCGCAGCAGCAGCAGTACGAGTGAACCTTGCGCAATGAAGCGCGTGTGTTGAA 8237
 Qy 463 nValAspMetSerArgGluMetValAsnMetIleile 475
 Db 8238 COTGATAGTCGGGTAAACGTGACCAACTCGTTCGTC 8274

RESULT 37

US-10-369-37512
 ; Sequence 37512, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ. ID NOS: 47374
 ; SEQ. ID NO. 37512
 ; LENGTH: 2037
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas fluorescens
 US-10-369-493-37512

Alignment Scores:

Pred. No.:	1.1e-05	Length:	2037
Score:	153.50	Matches:	126
Percent Similarity:	31.42%	Conservative:	71
Best Local Similarity:	20.10%	Mismatches:	205
Query Match:	5.94%	Indels:	225
DB:	15	Gaps:	30

US-10-009-823a-1 (1-502) x US-10-369-493-37512 (1-2037)

Qy 5 LeuPheLeGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
 Db 4 CTCAATATCGGAGTGTGGCGGTGGCGGAGTTCATCTCTGCGTGTGACGGCAAC 63
 Qy 25 AsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeu 44
 Db 64 AACATGCCAACGTGACGACCGCGGTATTCACGACGACCAACCGTG----- 111
 Qy 45 PheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAla----- 62
 Db 112 -----CAGGGCACCAAGTCTCGATTTCAG 135
 Qy 63 -----GlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 Db 136 TACGGTAACGTGTTTCATCGGTACGGTACGACTGTGCGACGTCGCGCGGTGTACAAC 195
 Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeu-----AlaIle 91
 Db 196 TCCTACCTCGAATCGCAGCTGCGTACCGCCACGTCGCTCAATAGCAGCGCGCTTC 255

Qy 92 GlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGly 111
 Db 256 GGTGCGCAGCGG-----ACCGCGGTGGACGCTGCTGCTCGACACCAACACTGCGC 306
 Qy 112 AsnPheAtgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131
 Db 307 CTGACCGGTGTGCTGCAGAAATTCCTCAC-----TCGATGCAAGGC 348
 Qy 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThrLeu-----GluPro 147
 Db 349 GTGTGACCTCGGCCACCGACGACACTTCGCCCGAGTCGGTGTGACCGCGGCCGAGGCC 408
 Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
 Db 409 CTGACCACTGCTTCAAT-----CGCTGGCCAGCAGATGAACGATCAGACGCCACG 462
 Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
 Db 463 CTCAATGGC-----AACCTGTCCGACATGGCGTCCAGTGAACAACTGCAACCTCG 516
 Qy 188 TyrPheAlaLeuGluSerTrpLysGlyAsnGlyThrProIleSerThrSerAsn 207
 Db 517 ---ATTGCCAGCTGAACACGAGATCGCGAG-----ATCTCCACCGCGGC 561
 Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAsp----- 218
 Db 562 -----GGCCAGCCGAGCATCTGCTGCACAGCCGTAACGAAGCGTGCCTCAATC 612
 Qy 219 -----GlnGlnGlyAsnSerHisAspIle----- 226
 Db 613 TCCGAGTGAACCGCGCGCAAGTCTTGAGCGTGACCACTTCGACATCTACATCGGC 672
 Qy 227 -----ThrValTyrPheAsp 231
 Db 673 AGCGGCCAGCGTGTGTCATCGGCAATACCAATACCTCGAGCAGCGTCCGCTCAAG 732
 Qy 232 GlyAlaProSer-----SerThrGlySerLysThrPheGlyTyrLeu 245
 Db 733 GGTGACCATCGGCATCGGCATCGATGATCGTGTGTCGACACCATCGACATCACC 792
 Qy 246 ValAlaMetAsnProSerGluAspGly----- 254
 Db 793 TCGGCGATGACCGGTGTGTAATCGCGGTCTGCTGACTTATCGCAAGAGTGTCCAT 852
 Qy 255 -----Ser 255
 Db 853 CTTGCACCTCAACAGAGTGGCGCGTGTGCACTGTGTCGCGACAGATCAACCGCCAG 912
 Qy 256 AlaAlaSerGlyThrAspSerAlaGly----- 264
 Db 913 CAAGCCAGGGATCGACACAAAGCGTGACTTTGGCGGGCCATTTCAACACATCAAC 972
 Qy 265 -----LeuLeuMetSerGlyThr----- 270
 Db 973 AGTGGCGCTCATCAGTCAGTCAGCGCAGCATGCCCGAGTCCGCAACAGTCAGGTTCGGGC 1032
 Qy 271 -----MetThrPheSerSerAsnGlyGluLeu----- 279
 Db 1033 AATCTTGATGTCACATCAAGACACCGGACAGCTGACACCGACGACTATCAGGTCAAC 1092
 Qy 280 ---LysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAlaTrp 298
 Db 1093 TTCACGACGGCGCAACTACACGCTCAAGCGTCCGACGCGCACCGACATGGGTTCGTC 1152
 Qy 299 Gln-----ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
 Db 1153 AGCACCAGCACCGCGCTCCGTTAATCGACGCTTT----- 1191
 Qy 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsnMet 334
 Db 1192 -----ACCTGGCCCTTAACCGTGTGCTGCTGAGCGCGCGGATACG 1233
 Qy 335 TrpAlaGlyAlaProAlaSerAlaAla-----IleGlyThrAspIle 349

Db 1234 TTCAGGTCACCCGACCGGCGGTCCAGCATTACAGCCGTCCTACGACCCG 1293
Qy 350 GlyLysLeuProSerMetMetProLeuGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
Db 1294 AAGAAATACCGCGCGCGGACCATG-----ACCGGGTGGCCAGTGCACCAAC 1344
Qy 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
Db 1345 TCAGGCACC-----TACACTCAGCCGACGCTGACTGAC 1377
Qy 390 Val-----ThrIleThrSerGluGlyLysLeuGlnGly-----LysTyrSer 403
Db 1378 GTCTCGATATCTACACCCGCGGTCAGAGTTCAGACCGGCTCAAGTATTTCG 1437
Qy 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
Db 1438 ACCCGGTCAAACTGGTATTC-----GGT 1461
Qy 424 LeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Db 1462 GCCACACAGCGGACGAGCTTACAC--ATGGTTGATGCCAAGGGGCCACGATC 1518
Qy 444 Gly-----LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db 1519 GGTTCGGGGTCATCGTTCCTCCGTCAGGCGAACCCTGAAGATCGGCATGGTC 1578
Qy 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle-----IleGln 477
Db 1579 GACTCCACCGGTCTCCGGTG-----ATGGACACACCGTCATCCCGAACGTGCNA 1629
Qy 478 ArgGlyPheGlnMetAsnSer 484
Db 1630 AAGACCTTCACCGTCGACACC 1650

RESULT 38
US-10-398-221-1330
; Sequence 1330, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398, 221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1330
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1330

Alignment Scores:
Pred. No.: 2,16e-06 Length: 554
Score: 152.50 Matches: 57
Percent Similarity: 37.76% Conservative: 34
Best Local Similarity: 23.65% Mismatches: 81
Query Match: 5.90% Indels: 70
DB: 16 Gaps: 7

US-10-009-823A-1 (1-502) x US-10-398-221-1330 (1-554)
Qy 216 ValTyrAspGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla---Pro 234
Db 37 GTATATGATATGT-GCGGAAACATAACTCAAGTGAATGATGAAGCTGTACACCA 95
Qy 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254

Db 96 GATCGGAGCGGTAAATGTTCTTACGAATATGAATTCAAATG-----GACGGA 143
Qy 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
Db 144 AAGCATTAATCTCTCCA-----GTAACAGAAACACTTAATTAACAT 185
Qy 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 186 GCGCAGCGCGAATTACAAACCCAGACGCA----- 215
Qy 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 216 CTTAAAAATATCCAAATCAATTCACAGTAAACGGC----- 251
Qy 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 252 -----AAACAAGTCAATATG 266
Qy 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 267 -----GGCTTAAACCTAAGTGGCTTAAACCAAC 293
Qy 355 MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr 374
Db 294 TAC-----GGAAACAACCAAGTA 311
Qy 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
Db 312 TTCACCAACTTCTGACGGTAAAGGGCTGCGACTGTAAAGATTATGCGATTACCGAT 371
Qy 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
Db 372 TCTGGCTATATGCAATGAGTTACTCAGATGGTACAGTTATCCAGTGGCCCACTTGGC 431
Qy 415 LeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAla 434
Db 432 GTGGCTACTTCTCCATGAAGACGGCTTAGTCAAAATGGGAACGGCAATATGTTCCA 491
Qy 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
Db 492 GGATTATCTCTGGCGATGCGATATACGGGCTTGTGCGCAAAATGGCGTGGCGGAATT 551
Qy 455 Ser 455
Db 552 AGC 554

RESULT 39
US-10-335-977-17
; Sequence 17, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...810
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-335-977-17

Alignment Scores:
Pred. No.: 4,74e-06 Length: 810
Score: 151.50 Matches: 58
Percent Similarity: 43.30% Conservative: 39
Best Local Similarity: 25.89% Mismatches: 90
Query Match: 5.86% Indels: 37
DB: 16 Gaps: 8

US-10-009-823A-1 (1-502) x US-10-335-977-17 (1-810)

Qy 6 PheleGlyAlaThrGly---MetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
Db 13 TATTATCGGCCCAACAGGGCAATGGCTACACAATTTAAACCGCTTGGATTAAACCTTAAC 72
Qy 25 AsnleAlaAsnAlaenThrleGlyTyTyLysGlnGlnValValPheGlnAsp--- 43
Db 73 AATTAGCAACCTTAACACCAACGGCTTTAAAGAGACGATCGGATTACAGCGGATTTT 132
Qy 44 -----LeuPheSerGlnAspLeuAlaile 51
Db 133 TTAAGGCTTTACCAAGAAATACCGAGAGCACTGCCCTTAGAAGATCAACCAAGCGAGC 192
Qy 52 GlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal 71
Db 193 GCGAAGPATCTCAACCGCAACCTCAATCGTGTCCTATTCTATCAGAAATCTATACGGAT 252
Qy 72 ArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaile 91
Db 253 AGGAGTCTT-----GCGCGGTTGAAGGGACGAATAACCCCTAGATTTCGCCCTA 303
Qy 92 GlyGlyLysGly---PhePheGlnValThrLeuGluAspLysValHisTyThrArgAla 110
Db 304 ACAAGCCCTAACCTCTATTTCGATACAGACTPAATGAGGGCGCTCCCTTATACAGAGAC 363
Qy 111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet 130
Db 364 GGGCAITTCAGCGTTGATAAGACGGCTTTTGGTAACCCCTTAATGTTTAAAGTGCTT 423
Qy 131 GlySerArgIleSerAsnAsn-----ProAsnleLysLys 142
Db 424 ---TCAGCTCCCGGTTTGAACGAAAAAGGAGGATCATGCTCATGCTCAACGCTGAAT 480
Qy 143 GluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAla 162
Db 481 GAAGTGATCAATGGTGGAAATCATTCTTTAGGAT-----AATGAAGCC 525
Qy 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182

Db 526 CAATTCAGCGGGCGCGCTTGGCTTTAGTGAGTTTACGGAACCTAAATAATCTTAAAAA 585
Qy 183 SerGluAlaAsnProTyPheAlaLeuGluSerTrpLysGlyAsnGlyThrProPro 202
Db 586 ATAGGGCAAAACCTTTAT-----ACCTATCAGGGCGGAAGGCGTTTCATCAA 630
Qy 203 IleSerThrSer 206
Db 631 GTCTCTGACTCT 642
RESULT 40
US-10-282-122A-28180
Sequence 28180, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28180
LENGTH: 6615
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28180

Alignment Scores:
Pred. No.: 9,59e-05 Length: 6615
Score: 151.50 Matches: 119
Percent Similarity: 31.12% Conservative: 64
Best Local Similarity: 20.24% Mismatches: 220
Query Match: 5.86% Indels: 185
DB: 16 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-282-122A-28180 (1-6615)

Qy 3 GlySerLeuPheleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22
Db 3 GlySerLeuPheleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22

Db 4372 GGCAATATCGGCATCGGGCTCACCGC-----ACCGGTGAGATCGGGTTCGGGAGTTC 4425
Qy 23 SerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGln 42
Db 4426 AACTCGGCAGCCACCAATCGGCTGTTCACCTCCGGTACGGAAACAGTAGGATCTTC 4485
Qy 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55
Db 4486 AACTCGGCAGCCGCAACCTGGGATCGGAACACCGGCGCAACCTCGGCATCGCA 4545
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 4546 AACTCGGCAGCTTCAACACCGGCTCGGG-----AACCGGCGAGCACCAACAG 4596
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 4597 -----GGCTGTTCACCGCGGCAAGTCACACCGGCTCGGCAACACCGGAGCATC 4650
Qy 95 -----GlyPheGlnVal----- 99
Db 4651 AACACCGGAGCATCAACACCGGAGCTTCAACACTGGGAGGACCAATACCGGAGTTC 4710
Qy 100 ThrLeuGluAsp-----LysValHisTyrThr 108
Db 4711 AACCTCGGCGATCACACACCGGAGCTTCACTCGGTGACTACACACCGGCTACTTC 4770
Qy 109 ArgAlaGlyAsnPheArgPhe----- 115
Db 4771 AACCGGCGGATCACTACACACCGGCTGTGGCCACACCGGCAACGTCAACACCGCGGTC 4830
Qy 116 -----ThrGlnAspGlyPhe----- 120
Db 4831 ATCTCGGCAATTACAGCAACGGGTTCTTCTGGGAGGTGACTTACCAGGGTTGATTGCG 4890
Qy 121 -----LeuAsnAspProSer 125
Db 4891 CTTTCCACACGATCACCATTCCGAAATCCCTACCGCTACGACTGAGTGTTCATC 4950
Qy 126 GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsn----- 139
Db 4951 GACATACCCATCACCGGCAACCGCTGTCGCCACCACCGCAACAGTTTACCATTCGCGT 5010
Qy 140 -----IleLysLysGluThrLeuGluPro 147
Db 5011 TTCCAGATACAGTCTTGCTTGCTCGGCGGTGCTTGTCAACGAGATGATCGGCCCC 5070
Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
Db 5071 ATCAGATCGATGTCAATCAAGTCATCGCCATCGCCATTCGCCCATTCAGCAAAACCATCAGC 5130
Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
Db 5131 ATGTTGGCACCGGCGCTTCGGCCGATCCCATCGGCATCAGCATCGGTGTTACCCCG 5190
Qy 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsn 207
Db 5191 GGTTC-----GGCAACTCGACC----- 5208
Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227
Db 5209 -----ACCGCCCGTCTCGGGTTCTTCCACCGCGCGCCCATGTATCGGC 5259
Qy 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Db 5260 TTCGGAACTTCGGCGCGGCAACATGTGGGCTCCGGAACTTC----- 5304
Qy 248 MetAsnProSerGluAspGlySerAlaAlaSerGly-----ThrAspSerAlaGlyLeuLeu 266
Db 5305 -----GGCGCTGGCAATTTCGGGCTTCTTTAACCGCGCGGCTTGGGC 5346
Qy 267 MetSerGlyThrMetThrPheSerSer-----AsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 5347 AATTCGGGCTACTGAATTCGGCGCGCTCGAGTCGGGTCTGGGAACTTCGGGCAACACC 5406

Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
Db 5407 ATCTCGGGGTCTACAAACAGCAGCAGCTGGAGATCTCGGACGCCGCC-----TTCGGCTCG 5463
Qy 306 GlyLeuProGlnPheSerAlaAsnPheValGlyValaGlyIleGln-----ProLeu 322
Db 5464 GGCATCGCAACATCGGCGCCAACTTCGGCGGCTGTCTCTCGCAACACCGGCAACCTG 5523
Qy 323 ThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
Db 5524 ACGTGAATTCGGGCTCGCAACACCGGCGGCTCAACCGCGGC----- 5568
Qy 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
Db 5569 -----ATCGGAACTTCGGGAGCTTCGCGAGCTCAACATCGGCTTCGTTAATACC 5610
Qy 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyr----- 377
Db 5611 GGCAGCTCAACACTCGGCGATCGGCAACCTCGGCGAOCCTCAACTTCGGCGGGTCAACATC 5670
Qy 378 -----SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
Db 5671 GCGGTAAACATCGGCATCGCCACACCGGATCTTCGATATCGGCTTGGGCAACCTG 5730
Qy 396 GlyLysLeuGlnGlyLysTyrSerAsn-----SerGlnValValAspPhe----- 410
Db 5731 GGCAGCTAACATCGGCTTGGCAATCTGGGCGACGACAACTTCGGGCTTGGCAACGCC 5790
Qy 411 -----TyrAsnIleProLeuAlaArgPheThrSerGluAsp----- 422
Db 5791 GGCAGCTAACATCGGCTTCGCAACTTCGGGAGCGCAACCTTCGGGCTTGGCAACACC 5850
Qy 423 -----GlyLeuArgGluGluGlyAsnAsnHisTyrSer----- 433
Db 5851 GGCAGCTAACATCGGCTTCGCAATACCGGTAAACAAACATCGGCGTGGGCTCACC 5910
Qy 434 -----AlaThrLeuAspSerGlyGlyProGluPheGlyLeu--- 445
Db 5911 GGCAACGGCCAGATCGGAGATCGGCGCTCAACTCGGCGAGCAACATCGGCTGTTTC 5970
Qy 446 ---ProGlyThrSerAsnTyrGly 452
Db 5971 AACTCGGCGAGGAAACATCGGG 5994

Search completed: October 31, 2004, 03:22:14
Job time : 2291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 03:11:41 ; Search time 4223 Seconds

(without alignments)

5621.461 Million cell updates/sec

Title: US-10-009-823a-1

Perfect score: 2586

Sequence: 1 MMSLSFIGATGKHTSTGIG.....NSKSVTTADTMIQKALELKR 502

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2/1/USFTO.spool/US1009823/runat_26102004.100129.1603/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPL=0 -ICPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosome2 -TRANS=human40.cdd -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1009823 @CGN 1.1 2527 @runat_26102004.100129.1603 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.btg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1235	47.8	300343	1	AE017310 Desulfovi
2	961	37.2	300029	1	AE017314 Desulfovi
3	693.5	26.8	110000	1	AE017180.04 Continuation (5 of
4	595.5	23.0	9612	1	AF122909 Treponema

C 5	595.5	23.0	123663	1	AE017255
C 6	579.5	22.4	349965	1	BX842655
C 7	567.5	21.9	2516	1	TEU04619
C 8	563	21.8	110000	1	CR522870_30
C 9	557.5	21.6	12559	1	AE013853
C 10	557.5	21.6	290155	1	AE017132
C 11	557.5	21.6	335050	1	AE017132
C 12	550	21.3	10029	1	AE011450
C 13	550	21.3	300594	1	AE017291
C 14	534.5	20.7	5230	1	TEU28219
C 15	534.5	20.7	26536	1	AE001244
C 16	533	20.6	313050	1	BX321857
C 17	528.5	20.4	10303	1	AE013102
C 18	527.5	20.4	25022	1	AE004539
C 19	509	19.7	34817	1	BBU43739
C 20	509	19.7	44380	1	AE001137
C 21	503	19.5	1499	1	BORFLGE
C 22	502	19.4	203050	1	AL646078
C 23	498	19.3	1320	1	STFLGE
C 24	498	19.3	20753	1	AE008751
C 25	498	19.3	254050	1	AL627269
C 26	498	19.3	301983	1	AE016840
C 27	498	19.3	306358	1	AE016759
C 28	493	19.1	9224	1	AE005317
C 29	493	19.1	22605	1	AP002555
C 30	491.5	19.0	110000	1	U00096_11
C 31	485	18.8	19549	1	AE001509
C 32	485	18.8	90250	2	AC095028
C 33	484.5	18.7	250050	1	AL591976
C 34	484.5	18.7	260050	1	AL596166
C 35	484.5	18.7	290242	1	AE017324
C 36	484.5	18.7	349980	6	AX417038
C 37	484.5	18.7	349980	6	AX417039
C 38	484.5	18.7	349980	6	AX417042
C 39	484.5	18.7	349980	6	AX641666
C 40	483.5	18.7	13374	1	AE015135
C 41	483.5	18.7	292504	1	AE016981
C 42	483	18.7	2550	1	HFU09549
C 43	482	18.6	110000	1	BX950851_19
C 44	482	18.6	346613	1	BX571661
C 45	481	18.6	8513	1	AE012299

ALIGNMENTS

RESULT 1	AE017310	300343 bp	DNA	linear	BCT 27-APR-2004
LOCUS	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough, section 2 of 12 of the complete genome.				
DEFINITION	2 of 12 of the complete genome.				
ACCESSION	AE017310	AE017285			
VERSION	AE017310.2	GI:46577667			
SOURCE	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough				
KEYWORDS	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough				
ORGANISM	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales; Desulfovibrionaceae; Desulfovibrio.				
REFERENCE	1 (bases 1 to 300343)				
AUTHORS	Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L., Radune, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.				
TITLE	The genome sequence of the anaerobic, sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough				
JOURNAL	Nat. Biotechnol. (2004) In press				
PUBMED	15077118				
REFERENCE	2 (bases 1 to 300343)				
AUTHORS	Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B.,				

Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., Davidsen, T.M., Zafar, N., Zhou, L., Radune, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.

Direct Submission

Submitted (18-MAR-2004) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
On Apr 27, 2004 this sequence version replaced gi:46448084.

FEATURES

source Location/Qualifiers

1..300343
/organism="Desulfococcus vulgaris subsp. vulgaris str. Hildenborough"

/mol_type="genomic DNA"

/strain="Hildenborough"

/sub_species="vulgaris"

/db_xref="taxon:882"

753..1217

/locus_tag="DVU0257"

753..1217

/locus_tag="DVU0257"

/note="identified by match to protein family HMM PF00583"

/codon_start=1

/transl_table=11

/product="acetyltransferase, GNAT family"

/protein_id="AAS94740.1"

/db_xref="GI:46448085"

/translation="MQSIPCKVPLTQGHYSAVEAVSRSVTHAFLRDLTLIRA
RLVSDFLPNTLVGVFVWGLAGPAGTASRKLMLFIAPYQWQGLRRLVESAIRHE
GVAFVNEQPRALATYRNGFEVGRSGRDMGMPYLLHWRHEHDDQK"

complement(1378..3351)

/locus_tag="DVU0258"

complement(1378..3351)

/locus_tag="DVU0258"

/note="identified by match to protein family HMM PF00072;

match to protein family HMM PF00512; match to protein

family HMM PF00785; match to protein family HMM PF02518;

match to protein family HMM TIGR00229"

/codon_start=1

/transl_table=11

/product="sensory box histidine kinase/response regulator"

/protein_id="AAS94741.1"

/db_xref="GI:46448086"

/translation="MSESLVDDSEGIKVLGITLADAGYRVTTAAGEEAMRRFAE
ERPDIVTDIKMPSMGLLESIKRNDSEVIMLTGHDMELATQSLKRDATDPT
KPINDDVLDLGRARERITMRRRLRYTENEDMVAQSRALVAERQALALQVMDG
IASGVQTLRLMQAGIFNELPCFVSHSRYLEIVATNDLYKERLGHKVGQPSWEVYK
GREGSGNCPVWVETGEGORNEILIDKHGOEIPVIVHTAPILNDGEIEMVLEIS
VDVSKVRLQDELRIETERYRQLEDEAPCYIVIDRSKLVVDANREFFHDFGEAARH
CHDIFARLACQCPVRSFDDGKPHQYETVTTGDSQANVLVWTSPLRDAAGNIT
EVMEMSTIDTQQLQRLQSLGLLSTAHGKGLITLADGQVYRGLSGLDNRAR
VRDSFEDIRLTLGRRLVMDLILYAKKRLDNLWVVVAERFATETAELEGKAVRGV
GQLEIAAGSEGTAEAGALSALVNLLENVAEACASRSPHSVTFVSGVDANHE
FSVRDDGVGMDRETKELFTLFFSSKGSAGTIGLGFVANQVQQHGHIAVASPECHG
TTFVSLRPLPDVAPCPAQNGQPEGEADDN"

complement(3612..4001)

/locus_tag="DVU0259"

complement(3612..4001)

/locus_tag="DVU0259"

/note="identified by similarity to SP:P33394; match to

protein family HMM PF00072"

/codon_start=1

/transl_table=11

/product="DNA-binding response regulator"

/protein_id="AAS94742.1"

/db_xref="GI:46448087"

/translation="MAYKILVDDDDFYIVKVMVDLSNDNGYATCSASDGEALDLRR
EKEDLVLDLEMEVNGPFRVKTWREDFRNVPIVISGLSGIHLAINVASLKKP
FDNKLLEIRQLENPEGRSAGEH"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/locus_tag="DVU0260"

/note="identified by match to protein family HMM PF00072"

/codon_start=1

/transl_table=11

/product="response regulator"

/protein_id="AAS94743.1"

/db_xref="GI:46448088"

/translation="NRVLIVDEMGMRFLRLTLFTGCGHEAVVARGOEGELARAREIR
PLHLVLDVMPGGGLVYRRLRDLADLAATPVILSGVKGSTFAHALELYDAAAGTL
PPPEAYVKPRPDEILALAEELTGCTACG"

complement(4440..5366)

/locus_tag="DVU0261"

complement(4440..5366)

/locus_tag="DVU0261"

/note="identified by match to protein family HMM PF00582"

/codon_start=1

/transl_table=11

/product="universal stress protein family"

/protein_id="AAS94744.1"

/db_xref="GI:46448089"

/translation="MEKLIIFATSPADNAAKVAFLAEPNKAFLVVEHLGVPSR
GFSQVRLIRGESEHHDADYRDVLEEMKOTYSQLETYSQPVVQVQVAGVPHTEI
LRFARQEGADLIIVMGANTDPPGASRVRSVIGNTQVRSKARCPVLIVNRCPTCW
NLFNIVFSTDFSKAASAFALKTAQIGAKLYLHACDLAANPMGVFAGQVEIER
RIEKAALMQERYYSKMGDFENYQVDIWEGTPTVYELLKYARARQADLIIVAAHTKEG
VEEABIGSTVQVWLRSACPVASVSKPEKAVV"

complement(5370..5924)

/locus_tag="DVU0262"

complement(5370..5924)

/locus_tag="DVU0262"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/evidence="not experimental"

/product="hypothetical protein"

/protein_id="AAS94745.1"

/db_xref="GI:46448090"

/translation="MDDIRLNGDGIILREMDGTTVADPLACLRNVTLDEGYSURGYFA
LLGVHPLVRMSPLPALREVERCAGCGCLTESFERLEGRVVELVGFDDAARVEY
HTLRGITGEAGSPARSEIRFVPLGCLLDMLHLGRARHVLFGAARVLECDTFLRL
FDITGLAWELGPOGSGSQCSLGR"

complement(5946..6326)

/locus_tag="DVU0263"

complement(5946..6326)

/locus_tag="DVU0263"

/note="identified by similarity to SP:P94690"

/codon_start=1

/transl_table=11

/product="acidic cytochrome c3"

/protein_id="AAS94746.1"

/db_xref="GI:46448091"

/translation="MFRIRIGTVLAFATALLALAVAFARTDTPRLEPAALKPOORPAA
VFDHDAENKAKLENCAVCHHDKGKIVPEGSEGTGADCHAKATKTHLVNAVH
RQCNCHKTSGKGTSCGCHVRD"

complement(6352..7671)

/locus_tag="DVU0264"

complement(6352..7671)

/locus_tag="DVU0264"

/note="identified by similarity to SP:P33393; match to
protein family HMM PF00037; match to protein family HMM
PF02754"

/codon_start=1

/transl_table=11

/product="ferredoxin, 4Fe-4S, putative"

/protein_id="AAS94747.1"

/db_xref="GI:46448092"

/translation="MGITERIEDAGLMRGIAALTPERIEKVVKVAATEGEGAGRLKVV
YETCMRGCAKACHYCIHSDGEPDYSVGMKEQTWMLREGGVSPDVIYDMAQIA
YTCMRCRRCHYCPIGIDTGYISLVRIICHLGTVTPRWQMTAISHAATNQMWVK
EDENVDLSLWCEDEARSEIPTMRIPEVEGADILYSVIAPKPKFTQLIYQAAVINDQ
AGVSWTMPATPGWNSDMAMFSGDSVMGHEIKPAHELAQRLVKELIVGECGHAFRS
IYDVGRLWLRKMPHPVHVSFEFWELEFDQKIKLAKRFEPEVTHDFCNIVRGRL
MDKUREVVHALCENNVEMTPEHNFNCCCAGGGVINCPCPPFKGVRVEGNRIKAEQLKA

gene	TGKTVIAPCHNCHGLEPDIHHYKLGMEHTKFLGDIIYDCMEKPGAV"		Db	51822 CGGACGGGGCGTCTTTCCATCGCCCGCCCAAGCGGACAGGGGGCATGACTCTGTCTTCAAC		51881		
	complement(7691..8350)							
	/locus tag="DVU0265"							
	complement(7691..8350)							
	/locus tag="DVU0265"							
CDS	/codon_start=1		Qy	174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu		193		
	/transl_table=1							
	/product="membrane protein, putative"							
	/protein_id="AAS94748.1"							
	/db_xref="GI:46448093"							
gene	/translation="MQAFYLLIGPGATAFAVFLGGCAWRLSTMGALARRKDMVAVAYNESKSLMSIMHWTPPTLGRNENPMTVATFAPHISIVILALPAPCHATMWDYAFGLVAKPRADVALITIVLAACGPPARRLSDPTVSVFTRPMDNLVLGLAARVLTGILAKYVNTQNLVAFTHVLUSGEAMLVALPFLRLSHAVSFVTRAVIGSEFGAVRECKD"		Qy	194 SerTyrLysGlyAsnGlyThrProPheLysSerThrSerAsnTyrSerTyrAlaGlnPro		213		
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
CDS	/notes="identified by Glimmer2; putative"		Db	51942 GCATGGGACGGCACCGCTACCGCCCGCTTGGGACATCGCGGAGCTTCTCGTCCACATCGCTCCCGGGGAGAC		52001		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla		233		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52002 ATGCGCGTCTATGACGCTGGGGGACACCCACCACTCACCCTCGCTTCGACATGCC		52061		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp		253		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52062 ACCACCGCGCGGCGGATACGTCGAGTTCCTCGTCCACATCGCTCCCGGGGAGAC		52121		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe		273		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52122 GCGGTGCGCGCGGGTAGCGCGGTGCGGGGTGCTCATGTCGGGAGCGTTCGCAATTC		52181		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys		293		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52182 GGGTCGTTCAGGGCAGTTCAGGACATCATGGCCCTTCAGCCCATCAGGCGGACCCCAAG		52241		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	294 AspLeuAsnAlaTatGlnProAlaProLeu--ValAsnGlyLeuProGlnPheSerAla		312		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52242 GACCTCGCAACTGGTGCCTCCGCGCAGCTGACGCCCTCTGGCAGCGCGCAGTTCAAGCC		52301		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln		332		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52302 ACCTTCGGCGAAGCGGGCGGCGGACAGTGGCGCTCGACCTTGGCATCACCACCC		52355		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	333 AsnMetThrAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeu		352		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52356 ACCGCATGGAACAATGACCCGCTTCGGCGGGGGTGGCGCTCGCAGCCGCGCTCTT		52415		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer		372		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52416 GCGGGGCGCCACCGCCCAACTGGCTGCCACGTCGACCCGCTTACAGGAGGAGTTTCG		52475		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle		392		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52476 TCGTCGACCTACATGAAGCAGCAGCGCTATGCGAAGCGCTACTCATGAACCTTGAGGTG		52535		
	/evidence=not experimental							
	/transl_table=1							
	/product="hypothetical protein"							
	/protein_id="AAS94749.1"							
gene	/db_xref="GI:46448094"		Qy	393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn		412		
	/translation="MKRKYIASVAPLEHLMQBEAQFAPDGSAPAAVVCVEEGAFSLRVNDIVNDITFEKATIPRTPDRLTALVQIDGWTFLAVDGPWEETVGFVMDITFPGS"							
	complement(8359..9585)							
	/locus tag="DVU0266"							
	complement(8359..9585)							
CDS	/notes="identified by Glimmer2; putative"		Db	52536 GCAACGATGGCATCGTCAGCGGCAAGTACAGCAACGGTACGTACGTCTGCTCTCGC		52595		
	/evidence=not experimental							

DEFINITION Desulfococcus vulgaris subsp. vulgaris str. Hildenborough, section 6 of 12 of the complete genome.

ACCESSION AE017314

VERSION AE017285

KEYWORDS AE017314.1 GI:46449255

SOURCE Desulfococcus vulgaris subsp. vulgaris str. Hildenborough

ORGANISM Bacteria; Proteobacteria; Deltaproteobacteria; Desulfococciales; Desulfococcaceae; Desulfococcus

REFERENCE 1 (bases 1 to 300229)

AUTHORS Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L., Radu, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.

TITLE The genome sequence of the anaerobic, sulfate-reducing bacterium Desulfococcus vulgaris Hildenborough

JOURNAL Nat. Biotechnol. (2004) In press

REFERENCE 2 (bases 1 to 300229)

AUTHORS Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L., Radu, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source 1. 300229

/organism="Desulfococcus vulgaris subsp. vulgaris str. Hildenborough"

/mol_type="genomic DNA"

/strain="Hildenborough"

/sub_species="vulgaris"

/db_xref="taxon:882"

complement(170..1270)

/locus_tag="DVU1429"

complement(170..1270)

/locus_tag="DVU1429"

/note="identified by match to protein family HMM PF06071; match to protein family HMM TIGR00092"

/codon_start=1

/transl_table=11

/product="GTP-binding protein"

/protein_id="AAS95907.1"

/db_xref="GI:46449256"

/translation="MALSIGIVLPVNGKSTLFNALTKAQNAPAAAPYPTTIPKNAK VAVPDKLALATVQPVQVILVATVDFIDAGLVGASGEGGLNPLANTRECAIL EVKCPDENITVQDAIDPLRDIETITELLADTQSAEKPERQKMAKGDOKA AEEVRLAEMHNPASPFEASDAPLAWREGLTAKKIYICANVDITGEE NAHVLRVLEASRGADVVKAKIIEELQGLSDEQASMLSGYIDESGLVSLRTG YHILGLASVITRGEKVRATIKQNKAPQAGVITDFERFIRAEVIAVDYVKHR SEARASAGVIRVEGKEVQMDGDVHFLFNV"

complement(1684..4290)

/locus_tag="DVU1430"

complement(1684..4290)

/locus_tag="DVU1430"

/note="identified by match to protein family HMM PF05193; match to protein family HMM PF05193"

/codon_start=1

/transl_table=11

/product="peptidase, M16 family"

/protein_id="AAS95908.1"

/db_xref="GI:46449257"

/translation="MRFAALFTCTIIGVMLATSVQAAQPADITLSNGLTVLAKDDR FPLASLRVHAGASAPETPEAGISLHLEWFKVQKPGQVARDVESVGYLNA

TSFDVTVLTPMSTQWKLQMDVLRDWAEPALDPAELESKEKDVVIAELQGEDSPDS RIFQSOAGTLKGTTYVERPIIGYRETIATADTMRAVIRKHYQPOQSMILLTVGVNDP AEVRAEABRLFGGLTNQNTTPPAIDARAFAGPVVNVHGHKPKYVLGVLPVGL KALQAQDMLSLQGGPTALLRTFKYERQVDDISVANYSEFVGMXYIYALDLA DKVETFWKELTTLMLAGLKADAFTEQEFERLNIEDGLYRSKETVAGLAKSGEFGFF SKPGQGANLYQAVRVEREQLEALIRDLWRPERLTVSVLLPEGTAKPAAATLKSTW DSGPQGANLYQAVRVEREQLEALIRDLWRPERLTVSVLLPEGTAKPAAATLKSTW GLAATASVLTGTLKHDATLEAFQSDRAASLGASAGRTTFLSLRPSRFGQDMFG LLHEVLTALAPDEVAKEKNOVASIRAREDOPLGAHRLHLPFLPGHSGYGYHUG QPTEVEGTRDDVAKAFWAQQAQPMVMSVAGSFDEAVLPKSLPAPSKPVSILAP AWTEPKALDULPERNOAHLLVFTVGLAHKUTPALELLQSVLQSGGSLFLFRMRK QGLGYTVAMNQSDLAGFVLYITGTEPGLEQAEAGFRKVIDQLHATALPDEELRG KNOGRGYREHRLGSRSEAMLTSGQYPLLFNREVIDKAEKLAPSDLERVATYL DMNKVRVVKVLP"

4484..6679

/locus_tag="DVU1431"

4484..6679

/locus_tag="DVU1431"

/note="identified by match to protein family HMM PF01627; match to protein family HMM PF01740"

/codon_start=1

/transl_table=11

/product="hpt domain protein/STAS domain protein"

/protein_id="AAS95909.1"

/db_xref="GI:46449258"

/translation="MTDOEKRTQLERFLGALRVLDADAGVDLLELVGGDGEHVAVL GRLSRITQWRGAGGLLNMAHEALVLEAVLDLARHEKPLDTVVITAEIGMDAF AARFKWRDSQDAGTAEVTAQVQLQGVQVQPADARCI SERHVVYRGFDVDAFT MGHDLRASRGELRVYLAFDVHDLQKQVSPGLLVLFQKSGDVVDARFTGVGG GGVNVGLFSLIDALLQVAFQIDASKRFAVDTDVAARGEAAWRDLPAADGASA WPEPDSRSELILLQVDEAMKRLREQSQMDPEVDMNSDTDEKQASQAGSFDEV VMEDTSDGQEAAMPGLSLTSDVADPVYQAAGAEAYATDAPSSQAGADAMPPLDA PGSYFFDDAPGVGLTDPVADLTTEAGSLDAGDDADDDDAASLLULLEHDF DAPLDDADAYAVEPHRIGLAEGTLDGAASHLSGLSIADSLDLPDDAPDDQSH LDPLSRGKPGKGGDMLDGFDDAAGDEAPVDTPEAPAPFAAGAGLADSLVDP DVMAVGAAGGLPHEDEAIAEAPAGTPTSATSPSPSPSPSPSPSPSPSPSPSPSP VREGSGSLTSLGELTIERSLAFRDALLEMDRYQNVRLDLSGVTEVDLTFQLLRA AVVTAQRGGCTLAGGVVAPVSETAHRAGLDQAQVRAKRAFEAVLGMA"

complement(6906..9500)

/locus_tag="DVU1432"

complement(6906..9500)

/locus_tag="DVU1432"

/note="identified by match to protein family HMM PF04055"

/codon_start=1

/transl_table=11

/product="radical SAM domain protein"

/protein_id="AAS95910.1"

/db_xref="GI:46449259"

/translation="MRELLSLPKPSRYTGIEGSHKREDVAIHCAALFAFPDMYEVG MSYLGKILYLINDREDIWAERVFTPCREAGAILRSHNAPLSTLESDDTPLAALHLVG FSITHELQVNYLMLDLGILPLRSADRGDDLAAPTIIMAGGGCTLAAPADPDLIM VLGESEVQVYIIDLAKABSEGSTPEYLEAASSIPGVVPAFFADDAQAGTAPLR QGDKYVRRITADMDNARYAPAAQLVPFGAVHNRLAIEIGGCTRGCEGCGAGMYERA RERSVQVQALLEDCLARTGTDVSLSTGSDFSALKTLFQIVDRCABEQVSVSLP SLRVGSDDDIMRERAGIRGTALPEAGSQQRDVINKGVTEEGMLHRLKFLFEGH WQQVQLFMGLGPTTETEDLADVLCKRVDAAGRVRLQVTAASVPVPPKPTFP QWEAKIGLETRARVGLRDARREKGLRWHPEPMSFLEIGFSRGRDLAPVESA REKALIFSSNMEGFDLAPLWEMHGLSPDEVTCARDIDAPLAWHLNFEQDQOSH RERKALIKTTDPDRCYSGCCVCTKAGPSRLDPLGVAAYANLNFQDQOSH EPTLDDNGQVVRHEAFNRPPKLSGDLVHKASQVRIYVTKEDMAYVSOLEQITFERA LRRAGLPAPFSGHPLPLLSFGALPVGASRAEMWPTIILRFTAFDEVRQLEGRD PKGMRVTFVEEQFSGKHQPVVEFTRYLALPEEMQPTFIDAWRTFAACEAFEWTRD TKGQRTGDIRPLFRDITTFDDGSLTLRLDWTDFYMSPLARAVPEAFPEALMTK LEQHFEDA"

9499..9690

/locus_tag="DVU1433"

9499..9690

/locus_tag="DVU1433"

/note="identified by Glimmer2; putative"

/codon_start=1

/evidence="not experimental"

/transl_table=11

/product="hypothetical protein"

Db 16766 CCGTGGACATCACCCCC---AACCGCCGCCCATCGGTACGGAGCGCTCGCTGCTCGG 16710
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAaGAsnGlySerSerSer 373
Db 16709 GGTCTGACCTTCTCGCAGCCGACCCAGTCAGACAGCAAGTACGCGGTTCTGTCATCG 16650
Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 16649 ACACAGTTTCCAGAAACGAGCGGTACACGTTCTGCGGTCTTGTGAGAACATCACTGTGAC 16590
Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 16589 CGTACGCGCATCAGCGAGGGAATATCCAAACGTTGTGACTCTGACCTCTATCAGTTC 16530
Qy 414 ProLeuAlaAaPheThrSerGlnAspGlyLysLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 16529 ACCCTCGTGACTTCCACAGCAAGCAGAACCTGCGCGGAGAGGCGGCAACCTCTTCAGC 16470
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn----- 450
Db 16469 GGCACACAGATTTCGGGACCTT-----CTGCCCGGCCCTGCCAACAGCAACGGC 16419
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 16418 CTCGGGGCATCAGCTCGAATTCTCGAGCAGTCCAAACGTCGACCTCGCGCCGGAATTC 16359
Qy 471 ValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAla 490
Db 16358 GTGAGAGTATCACCACGACAGAGGTTTCCAGTCCAAACAGCAAGGTCATCAACACCACT 16299
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 16298 GACACCATGCTTGAAGTGTGTGAACATGAAGGT 16263

RESULT 3
AE017180_04
WPCOMMENT

Sequence split into 39 fragments LOCUS AE017180 Accession AE017180

Fragment Name	Begin	End
AE017180_00	1	110000
AE017180_01	100001	210000
AE017180_02	200001	310000
AE017180_03	300001	410000
AE017180_04	400001	510000
AE017180_05	500001	610000
AE017180_06	600001	710000
AE017180_07	700001	810000
AE017180_08	800001	910000
AE017180_09	900001	1010000
AE017180_10	1000001	1110000
AE017180_11	1100001	1210000
AE017180_12	1200001	1310000
AE017180_13	1300001	1410000
AE017180_14	1400001	1510000
AE017180_15	1500001	1610000
AE017180_16	1600001	1710000
AE017180_17	1700001	1810000
AE017180_18	1800001	1910000
AE017180_19	1900001	2010000
AE017180_20	2000001	2110000
AE017180_21	2100001	2210000
AE017180_22	2200001	2310000
AE017180_23	2300001	2410000
AE017180_24	2400001	2510000
AE017180_25	2500001	2610000
AE017180_26	2600001	2710000
AE017180_27	2700001	2810000
AE017180_28	2800001	2910000
AE017180_29	2900001	3010000
AE017180_30	3000001	3110000
AE017180_31	3100001	3210000
AE017180_32	3200001	3310000
AE017180_33	3300001	3410000

AE017180_34 3400001 3510000
AE017180_35 3500001 3610000
AE017180_36 3600001 3710000
AE017180_37 3700001 3810000
AE017180_38 3800001 3814139
Continuation (5 of 39) of AE017180 from base 400001 (AE017180 Geobacter sulfurreducens P

Alignment Scores:
Pred. No.: 7,31e-34 Length: 110000
Score: 693.50 Matches: 175
Percent Similarity: 48.80% Conservative: 70
Best Local Similarity: 34.86% Mismatches: 166
Query Match: 26.82% Indels: 91
DB: 1 Gaps: 13

US-10-009-823A-1 (1-502) x AE017180_04 (1-110000)

Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 49491 GCACGTGTACACCGGATCAGCGGTCTCAACGCCAAGCGGAGGCCATGTCGTCATCGGC 49550
Qy 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp 43
Db 49551 AACAAATTTCCAAACGTCACACCATCGCTTTAAGCAGGCGCGGATGCTCTTCTCGGAC 49610
Qy 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 49611 GTCTCTTCAGCACCATCAGCGGGGGTTC-----CAGATCGGC 49649
Qy 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 49650 CGCGCGTCCAGATCCAGACAGTGCAGAAATCAGTTTCAACCCAGGCGTCTTCGAGAGCACC 49709
Qy 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
Db 49710 GAGAGCGTACCGACCTCGCCATCCAGGGCGATTCTCTTCGTGGTCCAGAACACACGAG 49769
Qy 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
Db 49770 GCGCGTACTATACCGCGCGCGCGCTTCTCTTCAATAGGACAAGACCTTGTGAAT 49829
Qy 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
Db 49830 CGGAGGAGTATCAGTTCAGTTCAGGTCAGGTCATCTCCCTCGTCGGGACTTTCGACGCG 49889
Qy 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLys 163
Db 49890 GTGCTCAGCCCATCGATCTG-----ACCAACTTTGCCACCATCTCCGCCGAG 49937
Qy 164 ThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSer 183
Db 49938 CAGACTTCCACCGTCAAGTTCTGTTGAACCTG---GACTTCCACCCAGACACACCCGACC 49994
Qy 184 GluAlaAsnProTyrPheAlaLeuGluSerTrpLysGlyAsnGlyThrProPhe 203
Db 49995 CTGGCG-----TGG-----GACCCCGCAAAACCCGTT 50021
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 50022 GCCACGTCCAACTACTCGACCCAGC-----CTGTGCGGTCTACGATTCCAGGGCAATGCC 50075
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 50076 CACACCGCCACGGTGTATTTC-----CGCAAGACCCGCCGACACGATGGGAC 50123
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 50124 TGGCAGGTATCTCTCCCGGANGCGCGGAGGCGGCGGCGGCGGACCACTACCCCC--- 50180
Qy 264 GlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThr 283
Db 50181 -----ATCGACGGGACCCCTTCATGTCACCCGAGCCCTC-----ACC 50222


```

/ product="unknown"
/ protein_id="AAD20618.1"
/ db_xref="GI:4426944"
/ translation="MIQVRLNGTKYWINPHQIETIECNPDVTLQMLSGKYVYIKERP
EEILEAVAYRRKIGIFKNEL"
gene      3793..4583
RBS
CDS
/ gene="motA"
/ translation="motA"
/ db_xref="GI:4426944"
/ protein_id="AAD20619.1"
/ db_xref="GI:4426945"
/ translation="MDIASFIIGFAGIAVFAAGILGGSAGGLIHPASPMFIFITGSSYM
CLFLTYPLSYTIGIFKVARVFKVTDYGEKALVORFVALSEKSRAGLLALEEIEDF
EDFMRGLRNVDGIDEGEATFISLMENELNQEERHTWISLVNAWATLAPGFMIGT
VIGLIGMLLDKDTALGPNNAVALVTFYGLMQLNLLVPIATKLMYQNNEVRSKE
MILEGVIGIAGDNPRILAQRLITYLTPSDRKAIEAEVLKD"
gene      4575..5304
RBS
CDS
/ gene="motB"
/ translation="motB"
/ db_xref="GI:4426944"
/ protein_id="AAD20620.1"
/ db_xref="GI:4426946"
/ translation="MARKKKHGPSAAGSGMLTYADMTIMLCFFVMLFEPSEVDVTO
LMASAGISDPTGCGLSVAGSLDLGNTISSPMSEKMGKMLASALKKAVSLFAPEI
KTKNIATVSDERGIYSLASDVFYFGPSAEINIAESRDTLNLIAOLFLSSQDLASVFER
VEGHTSDGVDPNWNKSNWELSSARAINILHSLTDFGQESRFSVAGYADTRPISND
TAGRAYNRVVDIILLDDAHF"
gene      5334..5892
RBS
CDS
/ gene="flil"
/ translation="flil"
/ db_xref="GI:4426947"
/ protein_id="AAD20621.1"
/ db_xref="GI:4426947"
/ translation="MADNDLMDDDDIQENMASSVDTKRAGLIPMLIKWVAIVLVAL
IFIVVVVIMPNIRDKGSHSISPVSEVRETRDVLQYQALGILLKVTADRIPATL
IVDVALGYNNKSTQOELSARKVEIDFLRSYFKGKTAELRQEKIKIEKHIND
NVLTKNKKIDVRFTQXDIVEQ"
gene      5892..6941
RBS
CDS
/ gene="flim"
/ translation="flim"
/ db_xref="GI:4426948"
/ protein_id="AAD20622.1"
/ db_xref="GI:4426948"
/ translation="NTEVLSODEIDOLLTAISSGDDTDFRAVNDTRKIKIYDKPRP
DKSKQMRVQWHTFPAFLTTSLSAQLRSMAHVHVAATVEQLTEYEEIRISIPPTI
LAINMDPLRASALHSDPSTFISIDRLFGGQSKVQRELTIESSVMGIVPRTI
LANRERAWTIVLIDPNTDNPOFQIVTFSENVLLVLTETKVEEGNMNCLP
YITLEPISLKSQFWSVRRASTQYAAAIKDKLSTVEDVMAEVGSLDVSIDVL
NLRAGDVRLPNVRVDPFKLTGVRPKFSQPGVGRKGLAVQILEKIEDISGDBFEF
LTSEGDELVE"
gene      6919..8115
RBS
/ gene="fliy"
/ translation="fliy"
/ db_xref="GI:4426949"
/ protein_id="AAD20623.1"
/ db_xref="GI:4426949"
/ translation="MSDGSISQDEIDALLSGVGGGIASAPAAAGDLDLAGFKKALL
QFTKENIPGSSNLSNTEKGSVSEPIEFRESFIRKVSFENVTATLIDFSGSMG
DHAFMGPPELAKIVSLVNHEDNVITIDMALSVISETIAQYVGTETUSYLERAGITGVA
SSPAESHVPKAMMLPFRNFYSITVNVKLDLSAYQLWEILSDVDVKIATAGDPD
SMQMAAGGQNMGMAMQNMGMQMGATOTGPAQMGFGSGMPQGSPPQCGNMQA
MANNMGMQMGASVQVOPFPPLQGFVSOEBOGNIGLIMDVYMEMTVELGRTRMKIK
EILGMBGHIIELDKLAGEPVDVLVNHKPIAKGEVVVIBESFGRITTEILSPAERISD
I"
RBS      8128..8130
CDS      8141..8800
/ note="orf10; similar to Treponema pallidum orf3 encoded
by GenBank Accession Number U36839"
/ codon_start=1
/ translation="unknown"
/ protein_id="AAD20624.1"
/ db_xref="GI:4426950"
/ translation="MSLKKLFFIIFLFSVTLFAENDGTSDDGLSPSSGILLDAG
KKGNIERNEDGALPAGNSAFDLNVAERTQSPISRLQVLVSLIIVICILAYVVLFLK
KSLFSFSDSPYLKSVASINIAOGKSHVITLGEKAYIVGTIDASINIMIGVEDKTV
DTNWLDAERRSSSPKQDFASMLSSVFKGSKNDVDVNFPAQRERLNKAQAQVPEEK
E"
gene      8787..9612

```

Alignment Scores:

Pred. No.:	9.51e-29	Length:	9612
Score:	595.50	Matches:	167
Percent Similarity:	46.15%	Conservative:	73
Best Local Similarity:	32.12%	Mismatches:	205
Query Match:	23.03%	Indels:	75
DB:	1	Gaps:	13

US-10-009-823A-1 (1-502) x AF122909 (1-9612)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
DB	2168	ATGATGAGATCTTATTTTCGGGTGTAACCGAATGCAAAATACACAAACAAGATCGAC	2227
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
DB	2228	GTTATCGGAATAACGTTGCCACGTAATAACACAGGTTTAAAGAGAGGAAGAGTAAC	2287
QY	41	PheGlnAspLeuPheSerGlnAspLeuAla	55
DB	2288	TTTCAAGACTTAATTCACAGCAATTAAAGCGGTGCACCCGTCCTACCGAAGAACTCGC	2347
QY	56	SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValaThrIlePhe	75
DB	2348	GGTGTAAACCCCAAGAGTCGGCTTCGGTATGATGTCGCAAGTATTGTACAACTCTT	2407
QY	76	ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly	95
DB	2408	ACTCAAGAGGCTTGCACCAACACAGGTGTCAACACGACCTTGCATTCAGGGAACGCG	2467
QY	96	PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe	115
DB	2468	TTTTTCATCTTAAGACGGAGAAAACCTTTTATACAGGCGCGTTCGTCGATC	2527
QY	116	ThrGlnAspGlyPheLeuAsnAspPro---SerGlyPheThrLeuMet-----	130
DB	2528	GATAAAGAGAACCTTGGTGACACCCGCAACGGAATGAGAGTCCAGGCTCGATCGCT	2587
QY	131	-----GlySerArgIleSerAsnProAsnIleLysLysGluThrLeuGlu	146
DB	2588	CAAGAAGCTGAGAGATTCAGGATATCATACATCGCGCCACGCGAA-----	2635

```
QY 147 ProIleGlnLeuAspPheAsnSerProThrValAlaLysSerProAlaLysThrSerThr 166
Db 2636 -----GATTTAATATTCCTATAGGTCAAAACCTGATGCAAGGCAACCA 2693
QY 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
Db 2684 AGTGTAACTATGCTGTGAACCTTATAGAGAGCTGCTGAATGCGCAAGGACAAAC 2743
QY 187 ProTyrPheAlaLeuLeuGluSerThrLysGlyAsnGlyThrProPheLysThrSer 206
Db 2744 CGGGCTCAAAATTTGGAATCTACTCG----- 2770
QY 207 AsnTyrSerThrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
Db 2771 -----TCTACCGCAATTTAGGTATATGACAGCTTCGGTGAACCCACGAGCTG 2818
QY 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
Db 2819 CAAATCGATTTTCCAGAGTTCGGGGGAAGTAAATGCT-----TGGCGGCAACCGTGA 2872
QY 247 AlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer 262
Db 2873 AATGTCGATCCGCAAAATCGAGCGCTACAGCTACCCGTGTAGGATTTGGAACA---ACC 2929
QY 263 AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMet 282
Db 2930 GACGGCGTTCAAAACAGCTTCATAGTCCGTTTGTACATAAGCGGACATCTGCTCCGTT 2989
QY 283 ThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaPro 302
Db 2990 ACC-----GACACTCGAGAAACGTAACG-----TCTCCC 3019
QY 303 LeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeu 322
Db 3020 GCAGGCAAGTTTGTGTTCAAATTTTCATCAATGTAGTAGCGCAACCTGATGAA--- 3076
QY 323 ThrLeuAspPheGlyLeuLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
Db 3077 -----GCCGGAGCTCCC---ACAAGA 3094
QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
Db 3095 CATACATTTCAGCTAACTTAGCGAA-----ATCGGAATCTCCAA 3136
QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyr 382
Db 3137 AACACCATTTACTGATCTCCGATAAGAGCACTACAAAGCGCTATGAAAGGACGATAT 3196
QY 383 ProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyr 402
Db 3197 ACTCTAGTTATCTTGAACCTTTAGAAATCGCAAGCGGTATATTACCGCGGTTTAT 3256
QY 403 SerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAsp 422
Db 3257 TCCACCGGGTAAAGCGGAGAAATAGGTAGATAGTATGCGCGGTTTTCGCAATCAGGCG 3316
QY 423 GlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyProGlu 442
Db 3317 GGTCTTGAAGAAAGCGGACAAAACACTATGTTTCAGTCCAAATACGCGGTATTCGCAAT 3376
QY 443 PheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSer 462
Db 3377 GTTTCACATCGGGAACCGGTAGTAGGAGCTACTTTATCGCGGAACACTCGAATGATG 3436
QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMet 482
Db 3437 AACGTAGATTTAACCGCAACCAATTTGTAGACATGATCGTTACTCAAAAGAGCTTCAGGCA 3496
QY 483 AsnSerLysSerValThrThrAlaAspThrMetLeuGlnGlyAlaLeuGluLeuLysAsn 502
Db 3497 GGTGCTAAACAATACAACTTCGGGATACATGCTTGAACACAGTCTTGAACGAAACGA 3556
```

RESULT 5
AE017255/c
LOCUS
DEFINITION
Treponema denticola ATCC 35405, section 10 of the complete genome.
ACCESSION
AE017255 AE017226
VERSION
AE017255.1 GI:41819448
KEYWORDS
SOURCE
ORGANISM
Treponema denticola ATCC 35405
Treponema denticola ATCC 35405
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE
1 (bases 1 to 123663)
Seshadri,R., Myers,G.S., Tettelin,H., Eisen,J.A., Heidelberg,J.F., Dodson,R.J., Davidson,T.M., DeBoy,R.T., Fouts,D.E., Haft,D.H., Selengut,J., Ren,Q., Brinkac,L.M., Madupu,R., Kolonay,J.F., Durkin,A.S., Daugherty,S.C., Shetty,J., Shvartsbeyn,A., Gebregeorgis,E., Geer,K., Tsegaye,G., Malek,J.A., Ayodeji,B., Shatsman,S., McLeod,M.P., Majs,D., Howell,J.K., Pal,S., Amin,A., Vashisth,P., McNeill,T.Z., Weinstein,G.M., Norris,S.J., Fraser,C.M. and Paulsen,I.T.
Comparison of the genome of the oral pathogen, Treponema denticola, with other spirochete genomes
2 (bases 1 to 123663)
Seshadri,R., Myers,G.S., Tettelin,H., Eisen,J.A., Heidelberg,J.F., Dodson,R.J., Davidson,T.M., DeBoy,R.T., Fouts,D.E., Haft,D.H., Selengut,J., Ren,Q., Brinkac,L.M., Madupu,R., Kolonay,J.F., Durkin,A.S., Daugherty,S.C., Shetty,J., Shvartsbeyn,A., Gebregeorgis,E., Geer,K., Tsegaye,G., Malek,J.A., Ayodeji,B., Shatsman,S., McLeod,M.P., Majs,D., Howell,J.K., Pal,S., Amin,A., Vashisth,P., McNeill,T.Z., Weinstein,G.M., Norris,S.J., Fraser,C.M. and Paulsen,I.T.
Direct Submission
Submitted (29-JAN-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Source
1. 123663
/organism="Treponema denticola ATCC 35405"
/mol_type="Genomic DNA"
/strain="ATCC 35405"
/db_xref="ATCC:35405"
/db_xref="taxon:243275"
120..1004
/locus_tag="TDE2679"
120..1004
/locus_tag="TDE2679"
/note="Identified by similarity to ONI:NTL01BH1587"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAI313196.1"
/db_xref="GI:41819449"
/translation="MKKATFICLSVLNLFSCGKKAAETGKSTRAESYMAESKSSSDSPKEDKESDAVNTASLDNIEKRLKGTPEFEDIKKTRETIENLVTKYQAVISQEDHQFQSNIRQTSIRIPKPNFLNLTIGIKLDNKNITVEDVTBFVYGLNLRKVKETQYTKILSQAKTVKIDLEQVQDLRSIEATIEGLRYLQKSVNYSLTNSMYQIINSSITARPSPFTKALNAVKEGILGFLSDIIGILYWLIFILIAVAVLTKRKNIKKKSLDN"
complement(1001..1624)
/locus_tag="TDE2680"
complement(1001..1624)
/locus_tag="TDE2680"
/note="Identified by similarity to SP:083709"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAI313197.1"
/db_xref="GI:41819450"
/translation="MTVIAIVLVLNLFIFFFYFKFKQFSDRASINNKNETQKLVAQAFQTDQSVTIMEDKIVNSVLKDLILLACKBEKSAARVYESLAARAKNAQPSGDPAFVKTKPVSNSVKTERKSIENKKNPIDNEPKIYTKQLAPKSGIVPDQATASKKQIEMAKKGLSIELIAEKVPLPTCEIILISMNT"
complement(1621..1959)

Db	103002	GTATCGGAATAACGTTGCCACGCTAATAACACAGGTTTTTAAGAGAGGAAGAGTAAC	102943
Qy	41	PheGlnAspLeuPheSerGlnAspLeuAla-----lleglySerThrGly	55
Db	102942	TTTTCAAGACTTAATTTTACAGCAATTAAGCGGTGACCGCTCTACCGCAAGAACTCGGC	102883
Qy	56	SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValargThrIlePhe	75
Db	102882	GGTGTAAACCCCAAGAAAGTCGGCTTGGGTATGATGTCGCAAGTATTGATCAATCTTT	102823
Qy	76	ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly	95
Db	102822	ACTCAAGGAGCCTTGCAGACCAACAGGTGTCAACACCGACTTGCAATTCAGGGAACCGC	102763
Qy	96	PhePheGlnValThrLeuGluAspLysValHisThrArgAlaGlyAsnPheargPhe	115
Db	102762	TTTTTCAATCTTAAAGACGAGAAAAAATCTTTTATACAGGCGCGTCTTTCGGTATC	102703
Qy	116	ThrGlnAspGlyPheLeuAsnAspPro---SerGlyPheThrLeuMet-----	130
Db	102702	GATAAAGAAGAACCTTGGTGAACCCCGCAACGGAATGAGATCCAGGCTGGATGGCT	102643
Qy	131	-----GlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu	146
Db	102642	GAAGAAGCTGAAGGATTCAGGATAATCAATACATCCGGCCAGACGGAA-----	102595
Qy	147	ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr	166
Db	102594	-----GATTAAATATCTCTATAGGTCAAAAACCTTGATGCAAGGCAACCA	102547
Qy	167	AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn	186
Db	102546	AGTGTAACCTATCTGTAACTTTGATATAAGAGGCTGCCTGAATTCGCCGAAGGAGCAAC	102487
Qy	187	ProTyrrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProIleSerThrSer	206
Db	102486	CGGGCTCAATTTTGGAACTCTACTTGG-----	102460
Qy	207	AsnTyrrSerTrpAlaGlnProMetArgValTyrrAspGlnGlnGlyAsnSerHisAspIle	226
Db	102459	-----TCTACCGAATTTAAGTATATGACAGCTTCGGTGAAGCCACCGAGCTG	102412
Qy	227	ThrValTyrrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrrLeuVal	246
Db	102411	CAATCGATTTTCCAGAGTTCCTCCGGGAGTAAATAGCT-----TGGCGGCAACCGTA	102358
Qy	247	AlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer	262
Db	102357	AATGTGATCCGACAAATCAGACGCTACAGCTACCCGTGTAGGAATTTGGAACA---ACC	102301
Qy	263	AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMet	282
Db	102300	GACGGCGTTCAAAACAGCTTCATAGTCCGTTTTCACAATAACGACATCTTGTCTCCGT	102241
Qy	283	ThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaPro	302
Db	102240	ACC-----GACACTGAGAAACGTAACG-----TCTCCC	102211
Qy	303	LeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeu	322
Db	102210	GCAGGCAAGCTTTTGGTTCAAATTTTATACATATGTAGTAGCGCAACCTGATGAA---	102154
Qy	323	ThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla	342
Db	102153	-----GCCGAGCTCCC---ACAAGA	102136
Qy	343	AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer	362
Db	102135	CATACATTTGACGTAACCTTAGCGGAA-----ATCGGAACCTTCCAA	102094
Qy	363	GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrrSerGlnAspGlyTyrr	382
Db	102093	AACACCATTAAGTCTCTCCGATAAGAGCACTACAAAGCGCTATGAAACGAGCGGATAT	102034
Qy	383	ProGlnGlyAspLeuValAspValThrIleThrSerGlnGlyLysLeuGlnGlyLysTyrr	402
Db	102033	ACTCTAGGTATCTTGAATACTTAGAATCCACCAAGCGGTATTATACCGCGGTAT	101974
Qy	403	SerAsnSerGlnValValAspPheTyrrAsnIleProLeuAlaAspPheThrSerGluAsp	422
Db	101973	TCCAAAGGGTGAAGCAGGAATAAGTCAGTACTATATGCGCGTTCGCAATCAGGCG	101914
Qy	423	GlyLeuArgArgGlyGlyAsnAsnHisTyrrSerAlaThrLeuAspSerGlyGlyProGlu	442
Db	101913	GGTCTTGAATAAGCGGACAAACACCTATGTCAGTCCCAATACTCGGTATTGCCAAT	101854
Qy	443	PheGlyLeuProGlyThrSerAsnTyrrGlyLysSerValAsnGlnLeuGluThrSer	462
Db	101853	GTTTCCCATCTCGGAACCGTAGGTAGGTAAAGGCTACTTATTCGCGGAACACTCGAATGAT	101794
Qy	463	AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet	482
Db	101793	AACGTAGATTTTAAACCGCAATTTGTAGACATGATCGTTACTCAAAAAGGCTTTCAGCA	101734
Qy	483	AsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuArg	502
Db	101733	GGTCTAAACAATAACAACCTTCGATACATGCTTGAACAGCTTGAACCTTGAACACGA	101674
RESULT 6			
BX842655/c			
LOCUS			
DEFINITION			
BX842655 BX842601			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			
CDS			
gene			

gene
/gene="murF"
/locus_tag="Bd3204"
complement(675..2096)
/gene="murF"
/locus_tag="Bd3204"
/EC number="6.3.2.10"
/function="UDP-N-acetylmuramyl pentapeptide synthase"
/note="InterPro: Cytoplasmic peptidoglycan synthetases
N-terminal
hypothetical protein"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
6-diaminopimelate-D-alanyl-D-alanine ligase"
/protein_id="CAE78024.1"
/db_xref="GI:39576795"
/translation="MRAMDQTIIVKVTGAKILSQKTNFAGITDTRAQLOQLFIAL
KGEAFDHNFLDKAVEGASGLLVHENAILEKDKATILLYPDTLKALQKLNWAR
HESGARIVGIGSGNKTTFKEFTAAALVGSMDVHYNKGFNNHWGVPFTLLQLPFKD
VAITEMGNHAGETTELVAHAEPDVVCTWVGRHMEFFCTIEKVAEAKETIEAAGD
QTVRIYLDLTQTHMYVKADKEPKARILTFESSEDPADVHLMISNMMSRQISIKGS
IGGARFARVQFQAQNLMLMAASVGLAVGTPAQVHGLPACTKNGRQLHHLK
SQAMIFDANPDSMKALIDNNKLITVPRKRVGVGQNRKELGSSASNLHEELGVW
GQAGFDKVFYIGDDADAFKAGLSAGYKNLALIEKDYKSSGODLAGFLKTGDIIVK
ASRGTKLERFVFCEPLDFAEQ"
complement(2099..3589)
/gene="murE"
/locus_tag="Bd3205"
complement(2099..3589)
/gene="murE"
/locus_tag="Bd3205"
/EC number="6.3.2.13"
/function="UDP-N-acetylmuramyl tripeptide synthase"
/note="InterPro: Cytoplasmic peptidoglycan synthetases
N-terminal
hypothetical protein"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramoylalanyl-D-glutamate--2,
6-diaminopimelate ligase"
/protein_id="CAE78025.1"
/db_xref="GI:39576796"
/translation="MKLQHLFSILGPENAFSHIEVTGVFNDRALVFGSVFAIRG
NKLDGHTFIPDAVAKGAALVDEKAKIEGEGYGVVQVPSREVLVDLASRYWDPG
QELFCVGTGTNGKTSVTYMTAILELHNGKIPGTGIVTNHLDGQVWPSEMTTPDFV
LOKRLRFRSAGALAVAMEVSSHALDQRRVDSVPFNVTIPTNLTRDLDYHNTMSYL
EAKORLPDILLWTKRCPFAIVNTADKFRGRVRADPAVLMTYGEKSDIRYEILKM
DFALTHEKVTYVGEVRLPMSGTHNVNMLAALGAGLSAGIPLNLCIEALDRFTGV
PGRLOSPNKNLSVFVDYHTPDALENVMLTKYRNLQSSQRIWTFGCGGBDK
GRPLMAEMALKSDEVTSDNRPEDPQTILODILAGVGGANKAKATIVDRKDAI
TQTLKRAQEGDVLIIAGKHEDYQITGTQKFPFSDVKVAEEALQGR"
3617..4297
/gene="estA"
/locus_tag="Bd3206"
3617..4297
/gene="estA"
/locus_tag="Bd3206"
/EC number="3.6.1.1"
/function="predicted esterase"
/note="InterPro: Phospholipase/Carboxylesterase
hypothetical protein"
/evidence=not experimental
/codon_start=1
/transl_table=11
/product="serine esterase, putative"
/protein_id="CAE78026.1"
/db_xref="GI:39576797"
/translation="MRQGLKHQCEIHNDAPWVIFPHGYGADANDFSLGELIPTK
KTYNLPNGNLEVPICPAWGTGAWWTIDMWEIOAEGEHRDFNDTPKMSKAYD
LAMEIIPQMKVPNNKIVLGGFSQCAMLATIYLIRAPETPKGLVIMSGTLVPHODEKQY
VFNPAQGRFFVQSHINDAVLGYKQAQKLETLITLQNGMKGSLQCGHGLIPHPVITQI
GEYLTITP"

gene
/locus_tag="Bd3207"
4339..6081
/locus_tag="Bd3207"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE78027.1"
/db_xref="GI:39576798"
/translation="MKTLITLAFLLTSFAFAHRLVLEDPSEIQTLSEADPNLGRVL
SLPGADNKKLLEHPAYNSLVADLEKLEALUKTQSKUGVGNITVHRLDFIRWLKSPAA
RFEVLGVNMDRAVFNPGTCGEVRLIYRLAYTKTQNTLIQRLPMTLNAVKFLPAT
PNCQAVAKQSSQKQPLKPWVTLANKLSLEVNLOAVRWFSTIGDMGHAEPFLRY
LONNVFVASAMENTPDVARSQNSALRAELLAWISTPENIKALDAGILNIPKFLPT
KSSSPALHGMNRLANRPDQIFKEDFNKLTDLENIYGPSFRRLNLDSCAGCHQ
QRTIAGFHFLGKDPSTIFANSVFSQS PHMOEQKEROVYFQNI FQKKAADSRPFS
ERDPTPGAMNAGCLFGSEYQAWTCAAGLKMAYVTPESKSEIGECLEPQWAGNPC
EPGVSTQANSHGCKMAGRTISCPSHQYQTTIKVFPFGMCGCGCLTSGEACGAI
AVLQGFNDCLARHPPFGECLAQNTRPAGLQACDDTISCADYTCNKTKQSGKVCIPPY
PLFQRLVDGHPKPI"
4339..4362
/locus_tag="sp0828"
4339..4362
/locus_tag="sp0828"
6108..6650
/locus_tag="Bd3208"
6108..6650
/locus_tag="Bd3208"
/locus_tag="Bd3208"
/function="cAMP-binding proteins - catabolite gene
activator and regulatory subunit of cAMP-dependent protein
kinases"
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="probable cyclic nucleotide binding protein"
/protein_id="CAE78028.1"
/db_xref="GI:39576799"
/translation="MRPHCLHHPHVAVPTTRILARNWQKSLIITYDPMFNASKASKKXP
IDQIVLKLADGEILFREGDLSREMYIVQKGAVEFVKRVEQTMLILGKVDGSMVGENS
LLESPLRSASAQAVGETTLLMDFDPSLLKRRDPTFAFELMKQLSGRIESTNEKLIT
LMAAQLSRNDLQIAESTL"
5647..7402
/locus_tag="Bd3209"
5647..7402
/locus_tag="Bd3209"
/function="Uncharacterized protein conserved in bacteria"
/note="Ser/Thr protein phosphatase family protein
hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative phosphatase protein"
/protein_id="CAE78029.1"
/db_xref="GI:39576800"
/translation="MIIDVTTEKMWISDLHLGNPFSEAKSLIEFLYAAENKYDYC
INGDGLTAQVSPRKAEDVPEVFRVAKAITKAGRVFYVVGNDHIVLENFLEDWGL
TLAPLNVRCGKORIRIEHGHLYDPFFVKRPDLVEFLTLGGFALWVSPRIYKAWIKP
BELSKRNRPRGSGSEVPCLPGEHPDVIAAEEICSRGFDALIFGHTHPTIDLPQG
RIFNSGSMWISPTTIGITISQDVLSQFEADKGA"
complement(7409..9001)
/gene="recQ"
/locus_tag="Bd3210"
complement(7409..9001)
/gene="recQ"
/locus_tag="Bd3210"
/EC number="3.6.1.1"
/function="superfamily II DNA helicase"
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=11

Db	181034	ATCACTGGCGGGAAGGCAAAATGTCGAAGTGGCTGCTAACTGATGTTCAACC	180975
Qy	275	SerAsnGlyLeuLeuLysAsnMet-----ThrAlaPheThrProThrGlySer	290
Db	180974	GTTGACGGTAAATCGATTCAGGAAACCACTCCAGAACTTCACTTCAAGGTGGT	180915
Qy	291	AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe	310
Db	180914	GCTTTGCGAGGAC-----	180903
Qy	311	SerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly-----Ile	328
Db	180902	-----CAGCAGTGAAGTGAACCTCGGTGATCGGATC	180870
Qy	329	LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp	348
Db	180869	AAG-----GAC	180864
Qy	349	IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg	368
Db	180863	GCGCGTAAA-----	180855
Qy	369	AsnGlySerSerSerThrArgArgTyrSer-----Gln	379
Db	180854	--GGTTGGACGGTACTAAGCAGTACGGTAAAACTCGGACCTGATCTCTTGGCACCAG	180798
Qy	380	AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlyGlyLysLeuGln	399
Db	180797	GATGCGCGGCTCGCGGTACATCACTGGCTTGTCTTCAATGACGAGGTACTTTGACA	180738
Qy	400	GlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThr	419
Db	180737	CGCGTATCTCCAAAGCTCAGGCCAATGACCTTGGCCAGATGGCTTTGGCAAGTTTGA	180678
Qy	420	SerGluAspGlyLeuArgArgGlyGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly	439
Db	180677	AAACCCAGAGGCTCTGTCTCAAGTTGGTAAACACCGTTTGAAGAAATCCAGAGACTCCGGT	180618
Qy	440	GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu	459
Db	180617	ACTGCTCCGAGTGTCTCAGGCGCTGCGCGCGGCGGTAACCTGTCGCAAAATCTTTG	180558
Qy	460	GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGly	479
Db	180557	GAAGAATCCACTGTCGATCTGGCTACAGAGTTCGTGAACATGATCCAGAATCAGCGTGGT	180498
Qy	480	PheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu	499
Db	180497	TTCCAGGCAATGCTCAAAACCATCACCAGCAGCGGCACTTCTGAACGAAGTATCCAG	180438
Qy	500	LeuLysArg 502	
Db	180437	TTGAACGA 180429	
RESULT 7			
TPU04619		2516 bp DNA linear BCT 27-JUL-1995	
LOCUS		Treponema phagedenis Kazan 5 flagellar hook polypeptide (figE)	
DEFINITION		gene, complete cds, MotA gene, partial cds.	
ACCESSION	U04619	GI:439979	
VERSION	U04619.1		
KEYWORDS		Treponema phagedenis	
SOURCE		Treponema phagedenis	
ORGANISM		Bacteria; Spirochaetales; Spirochaetaceae; Treponema.	
REFERENCE	1	(bases 1 to 2516)	
AUTHORS		Limberger, R.J., Sliwinski, L.L. and Samsonoff, W.A.	
TITLE		Genetic and biochemical analysis of the flagellar hook of Treponema	
JOURNAL	J. Bacteriol.	176 (12), 3631-3637 (1994)	
MEDLINE		94266716	
PUBMED		8206841	
REFERENCE	2	(bases 1 to 2516)	

/product="ATP-dependent DNA helicase RecQ"			
/protein_id="CAE78030.1"			
/db_xref="GI:39576801"			
/translation="MKVSFMEIIVSPQENOLKKYFNLSFRGQKEIIIESVMGGRDVL			
AVLPTGGGSKCYQPAVATVIVISPLIALMKDQVASLRRYIGIPAGALHSGSD			
DKREYFAINKGFAVLVLSPEAKQEGFHRWQNRVGLFAIDEAHCSQWHDFFE			
EVAQNLVLLKLPDVPDLTASATVLDLISHLKQKPERMVHGFRSNLYOYE			
LCEDEDAKLILLQSIKQTPERRIIVYCGTRKVTESIAAFLLQKFKGKSVGYHAGLTS			
Alignment Scores:			
Pred. No.:	7,378-26	Length:	349965
Score:	579.50	Matches:	160
Percent Similarity:	44.55%	Conservative:	73
Best Local Similarity:	30.59%	Mismatches:	171
Query Match:	22.41%	Indels:	119
DB:	1	Gaps:	16
US-10-009-823a-1 (1-502) x BX842655 (1-349965)			
Qy	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
Db	181703	ATTCTTTCTTCACTGATCGGTATCGGTATGACACAGCAGCGGCGAGGCTCTGGGT	181644
Qy	21	ThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValVal	40
Db	181643	GTTATCGGGGCAATATCCCAATGGAACATATCGGTTTCAAGCAAGCGGTCCAGAG	181584
Qy	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
Db	181583	TTCCAGGACATCATCTCCAAAATCTA-----AAAGGTATTGTGGTGGTAAAC	181536
Qy	61	GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPhe	80
Db	181535	CAAAATGGCGCGGTGTGAAGATCGGTGCGGTAAATCCATCTTCCAGGCTAACATC	181476
Qy	81	GluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThr	100
Db	181475	GACGCCACTGAAAAGTGAAGACCTTGCATCTCGGTGACGGTACTTCAAGGTAAA	181416
Qy	101	LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe	120
Db	181415	GGTTCCGCGGTGAGTCTACACTCGTACGCGGTCTCTCCACTTGCACCGTGAAGGCTAC	181356
Qy	121	Leu-----AsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnPro	138
Db	181355	CTGGTAAACAAACGAC-----AATCAACGCGGTACAGGCTTCTCT	181317
Qy	139	AsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAla	158
Db	181316	ACAGATGAGAGGCAATATCGTCAACAAAATGACCGATATCAAAATTCCTCGC---GCT	181260
Qy	159	LysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThr	178
Db	181259	CTGATCCCGGCTAAAGCAACCAAGAAATTTGAAACTGGATCTGAACCTGGATCCCGCATG	181200
Qy	179	AspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSerTrpLysGlyAsn	198
Db	181199	GACCCACCAAG-----AAGTTCGAC	181179
Qy	199	GlyThrProProLysThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAsp	218
Db	181178	CCAGCGGATCTTACTTCACTTCTCACTATCCACTGGT-----GTCAGATGATGAC	181125
Qy	219	GlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGly	238
Db	181124	TCTCAGGGAACAAACACCTTGCTGACGCTTCTTCTTC-----AACAAAGTTAAC	181077
Qy	239	SerLysThrPheGluTyr-----LeuValAlaMetAsnProSerGluAspGlySerAla	256
Db	181076	GACCGGAATGGGAATTCAAAGGTTTGGTT-----GACGGTAAAGAG	181035
Qy	257	AlaSerGlyThrAspSerAlaGlyLeu-----LeuMetSerGlyThrMetThrPheSer	274

AUTHORS Limberger, R.J.
 TITLE Direct Submission
 JOURNAL Submitted (20-DEC-1993) Ronald J. Limberger, Microbiology,
 Wadsworth Center for Labs and Research, David Axelrod Institute,
 New York State Dept. of Health, 120 New Scotland Ave, Albany, NY
 12201, USA

FEATURES
 source
 1. .2516
 /organism="Treponema phagedenis"
 /mol_type="genomic DNA"
 /strain="Kazan 5"
 /db_xref="taxon:162"
 /clone="phase 3"
 /clone_lib="gt 11 random partial digest"
 <1. .294
 /note="ORF1; homologous to ORF8 of Bacillus subtilis FlaA
 locus"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAA73466.1"
 /db_xref="GI:439980"
 /translation="PAPMEDTQFVAQMAOFTSLEOMTNVSQSFEKLNLLSGSEAVNA
 VGRQVNIEDGSIKVSQVITATRGRIPOVQVDCKWYDMSVTKTVIGEAQNR"
 319..324
 /note="flag site"
 334..1725
 /gene="flgB"
 334..1725
 /gene="flgE"
 /codon_start=1
 /transl_table=11
 /product="flagellar hook polypeptide"
 /protein_id="AAA73467.1"
 /db_xref="GI:439981"
 /translation="MRSIFSGVSGMQNHQTRMDVIGNVVNVNTTFKRGVNFQDL
 ISQOLGASPNVEGVGNPKVGLGVVVASIDVHTQALOTGTINTDIAIQNGGFF
 ILKDGKSFVTTAGAGVDRDGLVNPANGMVQGMWAEIDISQOIIINTSDQEDLII
 PIQKIDAKATTAVACHLDRPLPELPGANQADILRSTWADENVDTFGEQHLQ
 MVSFPGTNQNLATVDPENQAGTEIRVGIIDGTENTFIVSFNDYHGLASVTD
 TAGNVAPAGVLQVSYNVVANGPDEGGAPRHTFNINLGEISRNITITFABRST
 TRAYQDGYAMGYLEFKIDQSGIITGVYSGASREIGQALAGFANOGGLEKAGENT
 YIQSNNGIANITVSGVWGKGLIAGTLEMSNVLDLTQFTDMLITQRGFQAGAKTIQT
 SDTLETVLNLR"
 1740..1783
 /note="Mota gene site"
 2023..1027
 /notes="Mota gene site"
 2038..>2514
 /codon_start=1
 /transl_table=11
 /product="Mota"
 /protein_id="AAA73468.1"
 /db_xref="GI:439982"
 /translation="MDLASFIGFGAFAILMGILGGSAGFFHLPVSFIVFGSYL
 TLFAYPLSYTIGIFVKVCARVFSADPHEKEIVORLYALAEKSRRTGLLAEIEIQDF
 DDEFMRITGLRNVVDIGDAIRNLNEMELSHMEERHNRWISFINAWATLAPGVMLG"

stem_loop
 RBS
 CDS

Alignment Scores:
 Pred. No.: 1.28e-27 Length: 2516
 Score: 567.50 Matches: 160
 Percent Similarity: 43.95% Conservative: 69
 Best Local Similarity: 30.71% Mismatches: 215
 Query Match: 21.95% Indels: 77
 DB: 1 Gaps: 10

ORIGIN
 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 334 ATGATGAGATCAATATTTCCGCGGTTCGGAATGCAAAATCACCAGACAAGATGGAT 393

US-10-009-823A-1 (1-502) x TP004619 (1-2516)

21 ThrValSerAsnAsnAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 394 GTTATCGGAATAGCTGCACAACTGATACACAGGTTTAAAGCGCGGAGATTAAT 453
 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 454 TTCCAGGATTTAAATTTCTCAACAATTGAGCGGTGATCCCGTCCGAATGAAGAAGTCGGC 513
 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 514 GGTGTTAACCGAAGAAGTAGGCTCGTGTATGTCGCAACGATGATACCGGTACAC 573
 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
 574 ACACAAGGAGCTTGCAAACTACCGGAATCAATACTGATATTCGATACCAAGGAACGGA 633
 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 634 TTCITATTATTAAGACGAGAGAAAGAGTTTACACACAGCAGGTGATTCGGTGT 693
 116 ThrGlnAspGlyPheLeuAsnAspProSer-----Gly 126
 694 GACAGAGACGCGACATTTGGTAAATCTCTGCCAACGGAATGCGCGTGCAGGTTGGATGGCG 753
 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGluThrLeuGlu 146
 754 GAAGACATTTGAGGCGACCAATAATCAATACATCCGATCAAAACCGAAGACTTAATATT 813
 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
 814 CCGATTGCT-----CAAAAATTGATGCAAGGTGTACAC 849
 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
 850 GATGTTGCTGATGCGTGCACACTTGTATAAAAGTTTATCCCGAATTTGCCGGAAGGTGCAAT 909
 187 ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206
 910 CAACGGATATTTACATCACTG-----936
 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
 937 -----GCAACCGATTTTAACTATACGATCTTTCGGGGAACCAACATAAATCTG 984
 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
 985 CAGATGTTTCTCAAGGTCCTCCGCTGACAAACAC-----CAGTGGCTCGCA 1032
 247 AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeu 266
 1033 ACCGTAATGTTGATCCGAAATCAGGCCGGAACCGAAACCGGTGTAGGTATTGGAACA 1092
 267 MetSerGlyThr-----MetThrPheSerSerAsnGlyGluLeuLysAsn 281
 1093 ACCGACCGAAGCAAAATACCTTTATGTCAGCTTTGACAACTATGGGCAATTTGGGCTCG 1152
 282 MetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAla 301
 1153 GTTACC-----GACACGCGCGGAATGTTACA-----GCG 1182
 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
 1183 CCGCGCGGACAGGTTTATGTCAGGCTTCTTATAATGTTGTCGCGCA-----1230
 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSer 341
 1231 -----AACCTGATGAAGCGGTGCTCCCACT 1257
 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
 1258 CGTCATACCTTTAACTCAATCTTGGAGAA-----ATAGAACATCT 1299
 362 SerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGly 381

Db 1300 CGCAATACGATTACACAGTTTCCGAAAGAGCACTACAAAGCATATCAGCAGGACGGT 1359
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 1360 TAGCGATGGCTATCTTGGAATTCACAAATGACCAAGTGGTATTATCACCAGCGTA 1419
QY 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 1420 TATTCAACCGGTCAAGCCGTGAATCGGGCAGCTTGCTTGGCAGGTTTGGCAACACAG 1479
QY 422 AspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyPro 441
Db 1480 GCGGGTTTGAAGAAAGCAGGGGAAACACTTATTCAATCGAACACTCGGGTATTGCG 1539
QY 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 1540 AACATTACCGTATCAGGAGTTATGGTAAAGGAAAGCTTATTGACGGAACACTTGAATG 1599
QY 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGln 481
Db 1600 AGTAACGTGGAATTAACCGATCAATTTACCGACATGATTATTACAAAGAGGGTTTCAG 1659
QY 482 MetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1660 GCGGGAGCAAAACCACTTCAAACTTCCGATACATCTGGAAACTGTGTGATTGAAA 1719
QY 502 Arg 502
Db 1720 CGG 1722

RESULT 8

CR522870.30.

WPCOMMENT

Sequence split into 36 fragments LOCUS CR522870 Accession CR522870

Fragment Name	Begin	End
CR522870_00	1	110000
CR522870_01	100001	210000
CR522870_02	200001	310000
CR522870_03	300001	410000
CR522870_04	400001	510000
CR522870_05	500001	610000
CR522870_06	600001	710000
CR522870_07	700001	810000
CR522870_08	800001	910000
CR522870_09	900001	1010000
CR522870_10	1000001	1110000
CR522870_11	1100001	1210000
CR522870_12	1200001	1310000
CR522870_13	1300001	1410000
CR522870_14	1400001	1510000
CR522870_15	1500001	1610000
CR522870_16	1600001	1710000
CR522870_17	1700001	1810000
CR522870_18	1800001	1910000
CR522870_19	1900001	2010000
CR522870_20	2000001	2110000
CR522870_21	2100001	2210000
CR522870_22	2200001	2310000
CR522870_23	2300001	2410000
CR522870_24	2400001	2510000
CR522870_25	2500001	2610000
CR522870_26	2600001	2710000
CR522870_27	2700001	2810000
CR522870_28	2800001	2910000
CR522870_29	2900001	3010000
CR522870_30	3000001	3110000
CR522870_31	3100001	3210000
CR522870_32	3200001	3310000
CR522870_33	3300001	3410000
CR522870_34	3400001	3510000
CR522870_35	3500001	3523383

Continuation (31 of 36) of CR522870 from base 3000001 (CR522870 Desulfotalea psychrophil

Alignment Scores:

Score: 2.21e-25 Length: 110000
Matches: 563.00
Percent Similarity: 39.87% Conservative: 75
Best Local Similarity: 27.81% Mismatches: 190
Query Match: 21.77% Indels: 184
DB: 1 Gaps: 18
US-10-009-823A-1 (1-502) x CR522870_30 (1-110000)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 27852 GCATTATACACAGGAGTTCAGGGACTTAACACCAAGCCAGGCAATGACAGTATCGGC 27911
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp 43
Db 27912 AACAACTCGCAACACACACACTATTGCTTATAAGGATCTCGTACAATGTTTCAGAT 27971
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlnGlyProAsnGlnAlaGly 63
Db 27972 CTCATTATCCTCCAGCGTC-----ACTGGTTCGGGTGGCAGCTCTCAAGTAGGC 28019
QY 64 MetGlyAlaGlnValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 28020 CGTGGGTTGGCTCTCCGAGTGGACACAGATTTCAGTCAGGTACATTTGATCAACC 28079
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnVal-----ThrLeu 101
Db 28080 TCCTCCGAGCTTGATCTGGCCATTGAAGGAAACAGCTTTTCTCTCCAGCCCAAGCAAC 28139
QY 102 GluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeu 121
Db 28140 GACCAAGCATTCAATTACTCAGAGCAGGTTCCTTCAAAATTATAGGGTGGTACTCG 28199
QY 122 AsnAspProSerGlyPheThrLeuMetGly----- 131
Db 28200 GTAAACCCCGCGCTACGGGTTTCAGGAATAGCTACGGCGCAGAGCGGTGAACCTTGCA 28259
QY 131 ----- 131
Db 28260 GGTGGGATCTAGTATATTTTCTTGAGCGCTGGCATGATCAAGGTGGCGCCACC 28319
QY 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThr---LeuGluProIleGlnLeu 150
Db 28320 AGTGAATGATAATGAACACCAACTTTGACTCCAAACGCAAGATATTTCCCATCGACTC 28379
QY 151 AspPheAsnAsp----- 154
Db 28380 GATCCAGCAATACCAACTCCTATATTTTTCGAGGAGTCACTCCACGGGACCGCG 28439
QY 154 ----- 154
Db 28440 ACCGTTCTCGCACTGTCTCCGAGCAACAGCACCATCAGGCCCAACATATTACCGA 28499
QY 155 -----ProThrValAlaLysSer 160
Db 28500 AAAACCGCAGAGACAGCTGGAGTGCTACTACACGGCAAAAGAAACCATGGCAGACGT 28559
QY 161 ProAlaLysThrSerThrAla-----LeuAsnAlaVal 171
Db 28560 ACAGCTGGAGTTCAAGCTGCTTCCATCAGCGGGAGAGGTAGATAGATAGATCGCGCA 28619
QY 172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn---ProTyrPheAla 190
Db 28620 GGAACCTCGTTGACACAAAGCTCAGCCACGACCCCGAAAGCAAAATTTTGATCCCTTTGCC 28679
QY 191 LeuLeuGluSerTyrLysGlyAsnGly-----ThrProIleSerThrSer----- 206
Db 28680 ACCAGTATAGCTGGGACGACACCGGAGCCACCAACCGGTGTCTACGGGAGTACAGCG 28739
QY 206 ----- 206

Db	28740	GGAAGCCGTCGTGCTCAATAGACAAGAGCCTGGGGAACAAACAAACATCTGCACCTTT	28799
Qy	207	-----AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln	220
Db	28800	GACTACAGGATCCCAAGACCTTCACCTATTCGGCTCGGTACAAATTTTTCACAGCTTT	28859
Qy	221	GlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerThrGlySerLys	240
Db	28860	GGTAAACCGCATTAATACGACATATTC-----CGCAAAACAGACGACAAC	28907
Qy	241	ThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaIaSerGlyThr	260
Db	28908	ACCTGGGACACCTACTACTACTCGAA---AAGGAGGATGGCTCGCTCATTCCTGCAGAC	28964
Qy	261	AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys	280
Db	28965	CCCAACGCCATTTCCCTTACCAG-----TTACGCTTTGGCTCTGAGGCCAA-----	29012
Qy	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro	300
Db	29013	-----CCCGTTGACTCAGCG-----	29027
Qy	301	AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln	320
Db	29028	-----GGTCTCCCG-----	29036
Qy	321	ProLeuThrLeuAspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAla	340
Db	29037	-----CTTGAC-----GGCCCCCAACA	29054
Qy	341	SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr	360
Db	29055	GCAACCCAGGAGGCCATCGATTGGCGCAACCGCTCTGAGCCCAACGCCCTATC---ACC	29111
Qy	361	SerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArgTyrSerGlnAsp	380
Db	29112	GTTACCTTTGATACCCAGTAAACAGTAAACAGTAGCTCCAGGTTATCTCCTCAGAGCAGAC	29171
Qy	381	GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly	400
Db	29172	GGCTATGGCGGAGGAAACCTGAGCGGAGTAAATATCTGTCGCCAGGTCGCAATGTTATTC	29231
Qy	401	LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer	420
Db	29232	GCATCTCCCAACCGTAAACCCGTAACATTCGCGCAATCTCTCTGGCTAATTTCCAAAT	29291
Qy	421	GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly	440
Db	29292	CCTCGGGAGCTGACCATGGAGGGTAATACTTTTCTGCCACGGCCACCTCAGGAGCA	29351
Qy	441	ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu	460
Db	29352	CCAGGAGTGGTCTTACCCGGA---CCTGAGCTGGTAAAGTTTTCACAAATCTCGCTGGAA	29408
Qy	461	ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleLeuGlnArgGlyPhe	480
Db	29409	CAATCCATATGTCGATATGATGTCGATTTGTCGGATGATCACCATTCAAGCTGGCTAC	29468
Qy	481	GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu	500
Db	29469	CAGGCAAACTCTAAATATCATCACCACCGCTTGACGAACTACTCTGGCGAAATTAACCTT	29528
Qy	501	LysArg	502
Db	29529	AAACGC	29534
RESULT 9			
LOCUS	AE013853	12559 bp	DNA linear BCT 26-JUL-2002
DEFINITION	Yersinia pestis KIM section 253 of 415 of the complete genome.		
ACCESSION	AE013853 AE009952		
VERSION	AE013853.1 GI:21959364		
KEYWORDS			

SOURCE	Yersinia pestis KIM
ORGANISM	Yersinia pestis KIM
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.
TITLE	Genome Sequence of Yersinia pestis KIM
JOURNAL	J. Bacteriol. 184 (16), 4601-4611 (2002)
PUBMED	12142430
REFERENCE	2 (bases 1 to 12559)
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.
TITLE	Direct Submission
JOURNAL	Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES	Location/Qualifiers
source	1..12559
gene	/organism="Yersinia pestis KIM"
CDS	/mol_type="genomic DNA"
	/strain="KIM"
	/db_xref="taxon:187410"
	complement (87..2153)
	/gene="y2497"
	complement (87..2153)
	/gene="y2497"
	/function="putative"
	/note="residues 1 to 688 of 688 are 48.83 pct identical to residues 1 to 684 of 684 from GenPept : >dbj BAB53429.1 (AP03011) probable oxidoreductase [Mesorhizobium loti]"
	/codon_start=1
	/trans_table=11
	/product="putative oxidoreductase"
	/protein_id="AA086053.1"
	/db_xref="GI:21959365"
	/translation="MKSLSWDADAECVERLGLLGLPELGLRTYSARLLGSPKVLHGQNTSVKLNMDLFGNPQVMCIKSGQDLATITASGHPAVLAPLOQLKVLDDHLSIDMVNLTLLCRAPTSVEALLHAFVDRPVIDTHAVAMVITADQPDGRLCRYIYGERVWLPVMSGFALSVAVERAPRAYPDAGMLLKGCLCTFGSASAXSYERMITFVTLAEPIESRRATYSPVITPPAKPNLADMAHIAFYLRAPAFARSATSEKKHMTFDIRDPKSLAFSSGLGSDIAQRGVATPDHVRMKGRPVLDLADPLNLDGMAQOTQKIAAFERYQHYFHRNNARFERKIALDLPRAIVIFILGVGAGRTAEADAANADIIESLDVVTATVCGDYPIGEADFMEYWSLEQAKIDKASVKPQQQIYVVTGAGGGICAAIAKFEFSLGLAEALVDIDSESANKNAQLGFHALQCDVTEASVQAAPEMITAFKGVQDIVSNAGIALSGAIELEPTEATLRTSEFVNFFAHQVAAQAVSINKQGIQGLVLENIHQALNPGINFGAYCTSKAALLSLVROVALBQGDGIRVNAVNADRISGLLDDMISLRARAGLSSEKYMAGNLLGQEVTAQDVAKAFVYSMLDKXSTGNVITVDGNGVAAMLR"
gene	1395..1619
CDS	/gene="y2496"
	1395..1619
	/gene="y2496"
	/function="unknown"
	/codon_start=1
	/trans_table=11
	/product="hypothetical"
	/protein_id="AA086054.1"
	/db_xref="GI:21959366"
	/translation="WCHVSVQIVGLCGGRYHWRGIYGPAPFNKIFSQDGRNHSIRGFGLTKRTQAMLTQSHTLIRIVRPEGTDFRE"
gene	2371..3372
CDS	/gene="y2498"
	2371..3372
	/gene="y2498"
	/function="putative regulator"
	/note="residues 25 to 327 of 333 are 41.91 pct identical to residues 28 to 330 of 694 from GenPept :

```

QODGDVNVYSPAGGLSLVQGNPAKVEALPSSADATRLTLGVKRGNGEATEBVDSRIT
VLNKGTLKFRAEDLSARNQLGALWADSPFQTHNAGEDINDGEDEGFESPADPT
VTNKMDQGNVAITVEIKTKVKASDITVEFDIGDMIDTGGDITVETLSDNTKVQTTPGVNADG
DPTLEFGVAIDKINGTFPGAKDTFTIKTVSNVANINQVAILDSSKIAPAGSADGGI
SDNTNAQALLDQSKEKLTLSGAVAGLVSVNNGQATAKTNSTAAQAINIVTLTT
EQGISGVNLDEEYGDLQRFQQOYLANAQVLQAASTLFENALLISID"
complement (6912..7856)
/gene="flgJ"
/note="synonym: Y2502"
complement (6912..7856)
/gene="flgJ"
/function="putative structure; surface structures"
/notes="residues 1 to 305 of 314 are 58.44 pct identical to
        residues 2 to 305 of 313 from E. coli K12 : B1081;
        residues 7 to 314 of 314 are 61.29 pct identical to
        residues 8 to 316 of 316 from GenPept : >gb|AAL20112.1|
        (AB008751) flagellar biosynthesis [Salmonella typhimurium
        LT2]"
/codon_start=1
/transl_table=11
/product="flagellar biosynthesis protein"
/protein_id="AAM86059.1"
/db_xref="GI:21959371"
/translations="MSDLLANGAYDAQSLXEALEXDAARDPEGNLKOVACQVGMEVF
OMKLMRMALPDQGVNNSEQNKLVTSLDYDQIQAOQMGAQGLGLADMVQELSGSTSA
SETAGTVPMLDNELVMPAQALQVMRRAITPTPSSMAAISPGNCVFWARMSIPA
QIASQSGIPHOILINAALSEGWGOREIPTADGKSYNVFQIKAGSWNGPVSEITT
TEYEGVAKTKARFYGVGSYEAVSDYVKLLITONPRVARHVAQAQSPQGAAHLQKAG
YATDPYAOKLVSVIQMQESTGEQVAKVAGSGSDLQSF"
complement (7856..8965)
/gene="flgI"
/note="synonym: Y2503"
complement (7856..8965)
/gene="flgI"
/function="putative structure; surface structures"
/notes="residues 9 to 369 of 369 are 78.94 pct identical to
        residues 5 to 365 of 365 from E. coli K12 : B1080;
        residues 12 to 369 of 369 are 79.32 pct identical to
        residues 8 to 365 of 365 from GenPept : >gb|AAL2011.1|
        (AE008751) putative flagella basal body protein
        [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="homolog of Salmonella P-ring of flagella basal
body"
/protein_id="AAM86060.1"
/db_xref="GI:21959372"
/translations="MRKQSLVILLMVLVLSWLWPASABRIERLDVTQGVDRNALGYG
LVVLGDSGGQDTWTFTFTQSLSNMLSQLGITVPPCTNNQLKNVAAMVMTAKLPAFSR

```

Comment Scores:

No.	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3_85e-26	12559					
557.50	145					
43.98%	74					
29.12%	196					
21.56%	83					
1	11					

-0.009-822A-1 (1-502) x AE013853 (1-12559)

10 ThrGlyMethylSerHisSerThrGlyLeuGlyThyValserAsnAsnIlealaenAla 29
:::|||||
12408 AGCGGTATGAACGCAGCATCCAGTACTAGATGTCGTAAACAATATCGCCAACG 12349
:::|||||
30 AsnThrileGlyTyrlsGlnGlnValvalPheGlnAspLeuPheSerGlnAspLeu 49
:::|||||
12348 GCAACCTCAGGTTTTAAAGCAGGTTCAGTTCTTTCGCTGATATTGTTGCG----- 12298
:::|||||
50 AlaileGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
:::|||||
12297 -----GGTTCT-----CACACCGGCATGCCGTAAGGTCGCG 12265
:::|||||
70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
:::|||||

[illegible]

/protein_id="AA561568.1"
/db_xref="GI:45436012"
/translation="MSGKCKSTRNKGKVPMDRDEKSLAHTRWCKYHIVFAPKVRRO
VFVREBRAIGSLRLKCEKKNVILAECCVDHLMLEIPKMSVSGFMGLAKGS
SLMYPQGDLEKKNRFRWCRGYVDVVGKNTARIQYIKHQLEEDKNGEQLSIPY
PGSPFTRGR"
/complement(1183..2274)
/genes="ompA1"
/locus_tag="YP1326"
/complement(1183..2274)
/genes="ompA1"
/locus_tag="YP1326"
/notes="highly similar to Yersinia pestis CO92 ompA and
Yersinia pestis KIM ompA. Contains OmpA-like transmembrane
domain (Interpro|IPR000498). Contains MotY protein
(Interpro|IPR01035), (GO:001539), (GO:0009288)). Contains
OmpA outer membrane protein (Interpro|IPR002368,
(GO:0016021)). Contains Bacterial outer membrane protein
(Interpro|IPR006664). Contains OmpA/MotB domain
(Interpro|IPR006665)."
/codon_start=1
/transl_table=11
/product="putative outer membrane porin A protein"
/protein_id="AA561569.1"
/db_xref="GI:45436013"
/translation="MFMIMMRKMKKTAIALAVLGVPAVAQAPKDNWYTGKGL
GMSQYQDSINNDGTFHDKQAGAFGYQANQVLGEMGYDLGRMPYKGDING
AFKAQGVQAALSKVQAQDLVDYTRGLGLVWRADAKSGFDGDLRASGHDTRGVSPV
ALCAEAWKATWMEYQWNNIGRETVGARPDNGLSVGVSRYFGQEDAAAPIVA
PTPAPIVDKTRFTLSKSVLFGFNKANLPEQQAQDOLYAQLSSIDPKGSVVVLG
FDRQAPQNALSORRDSVRDVLVSGIPADKITARGEQANPVTGNTCDNVKPR
AALECAPDRRVEIEVKGEVTPQQA"
/complement(2602..3108)
/genes="sula"
/locus_tag="YP1327"
/complement(2602..3108)
/genes="sula"
/locus_tag="YP1327"
/notes="highly similar to Yersinia pestis CO92 sula and
Yersinia pestis KIM sula. Contains Cell division inhibitor
Sula (Interpro|IPR004596, (GO:0019866))"
/codon_start=1
/transl_table=11
/product="putative cell division inhibitor"
/protein_id="AA561570.1"
/db_xref="GI:45436014"
/translation="MRTQSLKPYHANYHSLTNDSTPRVDAPDTSGLISEFVYSNQ
VVTQLLPILQSLKQSRWLLWTPQQLSRWLSKQSLPINKVQLRQINELSTVEA
MEKALLTGNYSVVLGWLPELTEDDRIRLRLAAKLGNAVGVNRPLNDTKVSGQCANTL
KIHSYLVH"
/complement(3250..3972)
/genes="tfoX"
/locus_tag="YP1328"
/complement(3250..3972)
/genes="tfoX"
/locus_tag="YP1328"
/notes="highly similar to Yersinia pestis CO92 YPO1437 and
Yersinia pestis KIM Y2733. Contains TfoX, N-terminal
(Interpro|IPR007076). Contains TfoX, C-terminal
(Interpro|IPR007077)."
/codon_start=1
/transl_table=11
/product="regulator of competence-specific genes"
/protein_id="AA561571.1"
/db_xref="GI:45436015"
/translation="MRIFALSCSSYYCYKGTGLVIRFTKGVVWKMIMSKRIQAQVR
FGFLQQTIKSQGGYGLLVNSIMFVSEGEDILRANQNIEDLFRASGMALIVAKR
GLPVLRYRVDTSLMWDGRLRLHVRQAYEGAKIDVLSKGVVRLKDLPLNSASLE
RLWVKYIKNATRLIEGAKCCYLKLRLRSLRSLGINILLAGAISGYHYAALPLMR
SELIENPEKIHHPAKAPLYETA"
/complement(3826..3885)
/genes="tfoX"
/locus_tag="YP1328"

misc_feature

/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4074..6212)
/locus_tag="YP1329"
/complement(4074..6212)
/locus_tag="YP1329"
/notes="highly similar to Yersinia pestis CO92 YPO1438 and
Yersinia pestis KIM Y2732. Contains Sugar transporter
superfamily (Interpro|IPR005829, (GO:0016020))"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AA561572.1"
/db_xref="GI:45436016"
/translation="MFTFVIDLRRYAYNSLLHARIFIALVGVTAVPWVIEQKLI
PLTGVAAALTDDELTLRLNLFILICFFVASSSIELLPFPYPPFFALGLTSLC
GFTLGLGQRYATIAFCALLIAYITMLGASMTPTWYQQPLLSIGAVWNTLVFGH
LIFIRPLQONARCYQQLAHLEAKANLPDPIESINDINQPLIDVAVNSTLVATLQ
AKASLVRLLKRGQRGTRTRLYVFAQDIHERASSSHFYQVLRKFRYSVDLFRF
ORLLSMOAKCNHLSQILMHQYQHDPRFERAFTFLDAALARELAHENTQOVKAL
SYLLKNRAIDAQLAGISEQVLAQDPTPDNHLSDDEITGKSDIKLISLHLPQSAL
FRAIRMSIVLCVGIAFIQFTLRHGWILLTSLFVQCPYSATRRRLAIRIITLAG
ILVGLPILYVFSEIGOUTLIVISGVLPFAFNVQYAHMTFTILLVLCFNLLGEGF
EVAAPRVFDLLGCAIAWAVSFTWPKRQPLFTVVKTFNANCYRLDAILAQYHQG
KNSLRYVARRDAHNCDAELASVTSNMAEPKNDKETQEAARLLCLNHTMISYISA
LGAREKLTNSTTLLNLLDAICYTEGLAKEQPDNARMASALAELSNRLQCHPEPES
RDLVLAQIGLLELPELSLNKRISQPA"
complement(4611..4670)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4689..4757)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4827..4895)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5718..5786)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5796..5864)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5883..5942)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5952..6011)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6102..6170)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6242..6288)
/locus_tag="YP1330"
complement(6242..6288)
/locus_tag="YP1330"
/note="highly similar to Yersinia pestis CO92 YPO1439 and
Yersinia pestis KIM Y2731. Contains Protein of unknown
function DUF307 (Interpro|IPR005185)"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AA561573.1"
/db_xref="GI:45436017"
/translation="MRTVLNVNLFVGLGFTTGLMLATLVSVLITFLPLTRSCWEI
TKLSPLPYGNEAIHVNELYPEKSNALFTAGGALINIVMFLIFGWLCLSHITAGIVOC
MTIIGIVGINTFKIAIALMPVGGRRVVSVEAQAQARVANAQORH"
complement(6311..6379)
/locus_tag="YP1330"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6413..6481)
/locus_tag="YP1330"

Alignment Scores:
Pred. No.:

1.59e-24 Length: 290155

Score:	557.50	Matches:	145
Percent Similarity:	43.98%	Conservative:	74
Best Local Similarity:	29.12%	Mismatches:	196
Query Match:	21.56%	Indels:	83
DB:	1	Gaps:	11
US-10-009-823A-1 (1-502) x AE017132 (1-290155)			
QY	10	ThrGlyMetThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29	
DB	290028	AGCGTATGAACGAGCAGTACAGTATGATCGTACGTAACAATACGCCAATCG 289969	
QY	30	AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49	
DB	289968	GCAACCTCAGGTTTAAAGCAGGTTCTTCTTCGCTGATATGTTTTCG 289918	
QY	50	AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69	
DB	289917	-----GGTTCT-----CAGACCGCATGGCGGTAAAGTTCGG 289885	
QY	70	SerValArgThrIlePheThrGlnGlnAlaPheGluProGlyAsnSerValThrAspLeu 89	
DB	289884	GGGATTACTCAAGATTTTAAATGACGCGTACCGCCACCACTAATCGCTTTGGATCTG 289825	
QY	90	AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106	
DB	289824	GCTATCAGCCAGACGGTTTTCCTG-----ATGCAAGACAGCAGCGCGCGCATTTAC 289771	
QY	107	TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126	
DB	289770	TATCGCGTAAACGCGCATTTTAACTCGATGAAACCGTAATATCGTCAATATGCAAGGT 289711	
QY	127	PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu 146	
DB	289710	CTAAATCTAACGGTTTCCCGGCAACAGGACACCAACGGTACCAAGGCGGCTAAC 289651	
QY	147	ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166	
DB	289650	CCGGTCCCGCTGTCTATTCCGAGGATATGATCTCC-----GCCAAGGCGAGCAGC 289600	
QY	167	AlaLeuAsnAlaValAlaValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186	
DB	289599	TCCGGCAATATGGTGGCAACCTGACCTTACCATGATTTATCGCGAAGCAGCTCA 289540	
QY	187	ProTyrPheAlaLeuLeuGlySerTyrLysGlyAsnGlyThrProPheThrSerThr 206	
DB	289539	CTTTT-----GATCCCGATACCCAGAT 289516	
QY	207	AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226	
DB	289515	ACCTACAGCTTTGTCAATAACATGACGACCTTTGATAGCTTGGTAAACCCGCAATGAATC 289456	
QY	227	ThrVal---TyrPheAspGlyAlaProSerThrGlySerLysThrPheGluTyrLeu 245	
DB	289455	AATGCTTTTATGTTAAACGTCTGAAGATCGCACCGATCGGAATACCTGGGACGCTTAC 289396	
QY	246	ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265	
DB	289395	ACCAGGACAGTACGCGCAAGTACCGACCGACCGACCGACCGATCTGCTGCTGCC 289336	
QY	266	LeuMetSerGlyThrMetThrPheSerSerAsnGlyGlnLeuLysAsnMetThrAlaPhe 285	
DB	289335	GCTAAACGTGTTGATGTTTGTATGATTAACCGTCCCTGAAACCGTGAACACCGT 289276	
QY	286	ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsn 305	
DB	289275	ACA---AATGCAACAGTACACAGACTTCCAGTTTACGATCCCGATGGCGGTGTTAT 289219	
QY	306	GlyLeuPro-----GlnPheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThr 323	
DB	289218	GGTGGCGCGCACAAAGTTTTCGCACTGAATGTTGCGGGCAGT----- 289177	
QY	324	LeuAspPheGlyIleLysSerGlnGlnAsnMetTyrPheAlaGlyAlaProAlaSerAlaAla 343	
DB	289176	-----AAACAGCAAAAT----- 289165	
QY	344	AlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGly 363	
DB	289165	----- 289165	
QY	364	AsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrPro 383	
DB	289164	-----ACCGTGTGTCACAGCATTTGTGCACAGAACCCAGACTGGGTACGCC 289120	
QY	384	GlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSer 403	
DB	289119	CGCGTGAATTAATCTAGCTTCCAGATCAACAGTATGATGTTCTGTGGTAGGACTTATTC 289060	
QY	404	AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423	
DB	289059	AACACGACAGCAACCAATTCCTCGGCCAAATCGTCATGTTACTTCTTAACCCAGAGGG 289000	
QY	424	LeuArgArgGluGluGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 443	
DB	288939	GGGACTCGCGGTAGCGGTGCTCGGCACGTTGACAGTGTGCGCTGGAATCTCTAAC 288880	
QY	464	GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463	
DB	288939	GGGACTCGCGGTAGCGGTGCTCGGCACGTTGACAGTGTGCGCTGGAATCTCTAAC 288880	
QY	464	ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483	
DB	288879	GTGGATTGTGTAAGAGATGTTGTAACATGATTGTGGCAACGTAATATCATCTCCCAAC 288820	
QY	484	SerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501	
DB	288819	GGCGACAGCATCAAAACCAAGATCAGATCTTCAAACTGTGTTAGCTGCGT 288766	
RESULT 11			
LOCUS	AJ414150	335050 bp	DNA linear BCT 21-NOV-2002
DEFINITION	Yersinia pestis strain CO92 complete genome; segment 10/20.		
ACCESSION	AJ414150	AL590842	
VERSION	AJ414150.1	GI:15979723	
KEYWORDS	Yersinia pestis CO92		
SOURCE	Yersinia pestis CO92		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.		
REFERENCE	1 (bases 1 to 335050)		
AUTHORS	Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G., Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L., Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M., Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Karyshev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrall,B.G.		
TITLE	Genome sequence of Yersinia pestis, the causative agent of plague		
JOURNAL	Nature 413 (6855), 523-527 (2001)		
MEDLINE	21470413		
PUBMED	11586360		
REFERENCE	2 (bases 1 to 335050)		
AUTHORS	Parkhill,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk		
COMMENT	Notes: Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).		
FEATURES	source		
	1. 335050		
	/organism="Yersinia pestis CO92"		
	/mol_type="genomic DNA"		
	/strain="CO92"		

```

/db_xref="taxon:214092"
/notes="biovar: Orientalis"
147. .536
/gene="YPO1693"
CDS
147. .536
/gene="YPO1693"
/notes="Similar to Streptomyces coelicolor hypothetical
protein SC64.19C TR:086782 (EMBL:AL031317) (190 aa) fasta
scores: E(): 6.5e-10, 41.3% id in 104 aa, and to Bacillus
halodurans BH0898 protein TR:Q9KEF4 (EMBL:AP001510) (134
aa) fasta scores: E(): 0.034, 30.5% id in 95 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC90513.1"
/db_xref="GI:15979724"
/db_xref="EMBL:Q8ZFK0"
/translation="WATLPQKDYVERWAFGTEQVADDELLALVLNGTKTCAALDD
EGVPAQGDIFVVGNGRNPACAVELTEVELKTFDQDEHAHALEGEDSTLAYWRKTO
QRFERYDMFSDDMLCKMKFVLETF"
complement (646. .885)
/gene="YPO1694"
CDS
complement (646. .885)
/notes="YPO1694"
/notes="Similar to Escherichia coli hypothetical protein
YebV SW:YEBV ECOLI (P76274) (78 aa) fasta scores: E():
4.1e-19, 57.9% id in 76 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC90514.1"
/db_xref="GI:15979725"
/db_xref="EMBL:Q8ZFK9"
/translation="WKSIVKIQEVDVDAELSSOTKGEHTLISIPCKSDPDLQMLDG
WENSIPLALDGLTKLLYKQYHQDQDANWKVT"
complement (1093. .2082)
/gene="YPO1695"
CDS
complement (1093. .2082)
/notes="YPO1695"
/notes="Similar to Pseudomonas aeruginosa hypothetical
protein TR:AG08040 (EMBL:AE004879) (315 aa) fasta scores:
E(): 0.415% id in 299 aa, and to Pseudomonas fluorescens
hypothetical 23.6 kDa protein TR:Q9KIM7 (EMBL:AF228767)
(231 aa) fasta scores: E(): 1.2e-13, 29.9% id in 231 aa"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAC90515.1"
/db_xref="GI:15979726"
/db_xref="EMBL:Q8ZFK8"
/translation="MLSIITREITTTTPKYLWSWRTLPILWLVGNTLACTFSPT
VILPPNSFNVTLETSGSTIAMLCGTGVSLLASPVLNATVASTTTLKLNAGD
VIPYTIYSDAGYNNAIAGNVINAGNSLLIVTLFOISANLPVIRTSNNAINNV
SAGTYDVLNKNWYDLCAISFLSCLGGRFSGTATSTVTVKLTVSNDGVNSAPNV
FGSALITQAPNQNSATCKNATFTFTYITNGNVFVAPWPKMLTSGTDVLOYQLY
QHGAVIPWNTKRTATGTQQLIPYNAINVINPQQAQKTVGYQDNVSVVLEY"
complement (1663. .1728)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (1819. .1884)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (1945. .2010)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (2288. .4735)
/gene="YPO1696"
CDS
complement (2288. .4735)
/notes="Similar to Escherichia coli outer membrane usher
protein AfaC precursor SW:AFAC ECOLI (P53517) (859 aa)
fasta scores: E(): 1.4e-13, 22.7% id in 631 aa and
Pseudomonas aeruginosa hypothetical protein Pa4652
TR:AG08039 (EMBL:AE004879) (790 aa) fasta scores: E(): 0,
42.9% id in 805 aa"
/codon_start=1
/transl_table=11
/product="probable outer membrane usher protein"
/protein_id="CAC90516.1"
/db_xref="GI:15979727"
/db_xref="GOA:Q8ZFK7"
/db_xref="EMBL:Q8ZFK7"
/translation="MIIRFLPLSVLIYGSIIPLSLPILITLIGLSLPPSTSTTMDP
TTLYLELVNDRNFSTVPISYRNNRYLTSQSLRTIGLSEPLAEPALIDNAGYN
VKYDGENORLLINVPSEWLPKQLEVTQDDFNLAGSLGALFNVDIYATQGYPSYL
THPSAWEQRIFFDRGLSNTGVYRTHFRPSNNTDDAKGYRFDOTWKNDEHLLRY
STGDLITAGPMSAARLGGIOIARHAFIRPDILTYELPQFGQAAVPSIVDLYIDNF
RTQSANINPGFFVINNAPRINGAQATITVTDALGQISVSVFFVYASTILLKGVDF
SLSGALURRAYISANTSGAFNVAEPLNTNATNRLPLSPASRRGNRSLSGYSYS
GLGLVNLNAQIIISDDEYDLANKYTPSLSRMTQTLTSLGSLGSGYGVGFVDRDA
LGEOTRLINISYSTSLRNSFYALNRELGRKYNVQLVMSIPLGPRGSSISATRT
NDNQIOLNYSRAPSNGGLGNLAVANSTNNNOVQADIVRYTSMESRMGLYGN
SNYNYVAGLTSILVWRSVYASNMINDAFALVSTNGFSNI PVSYENQILGTNAKG
YLLIPTVASYIQAKFDPNLPADWMLPNVERLALGERSGYLINEPIKEISAVNIR
ITDASGDLPRGSAIYITGNIPISYVGWGNVIEQVAQNNLRIIRADNGTCYCSOF
KLKTGIDAGTTCVR"
complement (4923. .5675)
/gene="YPO1697"
CDS
complement (4923. .5675)
/notes="YPO1697"
/notes="Similar to Pseudomonas aeruginosa PA01 probable
pili assembly chaperone TR:AG08038 (EMBL:AE004879) (262
aa) fasta scores: E(): 0, 43.2% id in 243 aa, and to
Escherichia coli chaperone protein CscC precursor
SW:CSCI ECOLI (P53518) (232 aa) fasta scores: E(): 0.0013,
21.9% id in 242 aa"
/codon_start=1
/transl_table=11
/product="putative chaperone protein"
/protein_id="CAC90517.1"
/db_xref="GI:15979728"
/db_xref="EMBL:Q8ZFK5"
/translation="MIKYLHLLSKLLISFILLPIAPLHAASVLIWPIIDVINDENAT
ALMLEKDSQFTYQIRVLSRWQVNGLDNDAGTATTAIPSPVANIQKQQLIRLVKN
AQVPAKELAYRILIDIPQEPNDKNPSTGLGTFQMYRVSPLFVYNGWQIAPE
ATOPTDIAQPTSLYILNRQCKQWLTIRNQSEIHARISQVTLNGQTVNNGLLGYLAR
SOMSFPISLAARKGELAAIINSNTKPVITFYQ"
complement (5706. .6263)
/gene="YPO1698"
CDS
complement (5706. .6263)
/notes="YPO1698"
/notes="Similar to Myxococcus xanthus protein U precursor
pru SW:PRU MYXA (P27755) (179 aa) fasta scores: E():
6.7e-05, 25.0% id in 180 aa, and to Pseudomonas aeruginosa
PA01 hypothetical protein TR:AG08037 (EMBL:AE004879) (180
aa) fasta scores: E(): 8e-16, 34.5% id in 174 aa. This CDS
shares 31.9% and 30.8% identity with YPO1699 and YPO1700
respectively"
/codon_start=1
/transl_table=11
/product="putative exported protein"
/protein_id="CAC90518.1"
/db_xref="GI:15979729"
/db_xref="GOA:Q8ZFK5"
/db_xref="EMBL:Q8ZFK5"
/translation="MISLFPCHLSAALLISALSI VFTNAAPSPVSSPFTVNTVLTG
CVLGGGATDSNFGSINFGNITSLATINATISLNGTITIQCNQNPSTVLNLSGAN
MTGNISAGRHLLNSSTGEYLLIYIQYNSPRTLWGDGNGGATQVIATDGTLOQVILY
AQLFASNTQPTAGVYLDLIVTVTY"
complement (5976. .6002)
/gene="YPO1698"
misc_feature
/notes="PS00221 MIP family signature."

```

```

gene      complement(6284..6814)
CDS       /gene="ypol1699"
          /gene="ypol1699"
          /note="Similar to Myxococcus xanthus protein U precursor
          prU SW:PRU MYXA (P27755) (179 aa) fasta scores: E(1)
          9.9e-07, 28.9% id in 180 aa, and to Pseudomonas aeruginosa
          PA01 hypothetical protein TR:AG08036 (EMBL:AB004879) (177
          aa) fasta scores: E(1): 2.4e-20, 40.8% id in 170 aa. This
          CDS shares 31.9% and 35.9% identity with ypol1698 and
          ypol1700 respectively"
          /codon_start=1
          /transl_table=11
          /product="putative exported protein"
          /protein_id="CAC90519.1"
          /db_xref="GI:15979730"
          /translation="MKNTLVVPSLSLVGGLLFQAYPVNNAVKNTANITVTATLPTC

Alignment Scores:
Pred. No.: 1.88e-24      Length: 335050
Score: 557.50           Matches: 145
Percent Similarity: 43.98% Conservative: 74
Best Local Similarity: 29.12% Mismatches: 196
Query Match: 21.56%      Indels: 83
DB: 1                    Gaps: 11

US-10-009-823A-1 (1-502) x AJ414150 (1-335050)
QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnLeuAlaAsnAla 29
DB 115127 AGCGGTATGAACGACGACGATCAGTATCTAGATGTGATGATCAATATCGCAACTCG 115186
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
DB 115187 GCACCTCAGGTTTAAAGCAGGTTTCAGTTCTTCGCTGATATGTTGCG----- 115237
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
DB 115238 -----GGTTCT-----CAGACCGCATGGCGGTAAGTTCGCG 115270
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
DB 115271 GGGATTACTCAGATTTTAAAGCGGTACGGCAGCACCACTAATCTGCTTTGGATCTG 115330
QY 90 AlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106
DB 115331 GCTATCAGCAGAACGCTTTTTCGTT-----ATGCAAGACAGCAGCGCGGCATTAC 115384
QY 107 TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126
DB 115385 TATGCCCGTAAACGGCCAGTTTAAAGCTGGATGAAACCGTAATATCTCAATATGCAAGGT 115444
QY 127 PheThrLeuMetGlySerArgIleSerAsnProAsnIleLysLysGluThrLeuGlu 146
DB 115445 CTAATCTAACGGTTACCGGCAACAGGACACCAACCGGTACACAGGCGGCTAAC 115504
QY 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
DB 115505 CCGGTCCCTGTCTATTCCGACGAGTATGATCTCC-----GCCAAGGCGACGACC 115555
QY 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
DB 115556 TCGGGCAATATGGTGGCAACCTGACCTCTACCATGATGTTATCGGAGGACCTCA 115615
QY 187 ProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProIleSerThrSer 206
DB 115616 CCCTTT-----GATCCCGATAACCCAGAT 115639
QY 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
DB 115640 ACCTACAGCTTTGTCAATAACATGACGACCTTTGATAGTGTGGTAAACCCGATGAATC 115699

```

```

QY 227 ThrVal---TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
DB 115700 AATGCTCTTTTATGTTAAACGTGCTGAAGATCGGACCGATGGGAATACCTGGAGCTCTAC 115759
QY 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
DB 115760 ACCAGGACAGTACGAGCCCAAGTACGACGACGACGACGACGACGACGACGACGACGAC 115819
QY 266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
DB 115820 GCTAAACGTGTTTCGATGTTTTCGATGATGATGATGATGATGATGATGATGATGATGAT 115879
QY 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
DB 115880 ACA---AATGCAACAGTACACAGACATTCACGTTTACGATCCCGATGGGCGTGTAAAT 115936
QY 306 GlyLeuPro-----GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThr 323
DB 115937 GGTGCGCCCGCACAAAGTTTCGCACTGAATGTTGCGGGCAGT----- 115978
QY 324 LeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 343
DB 115979 -----AACACACAAAAT----- 115990
QY 344 AlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGly 363
DB 115990 ----- 115990
QY 364 AsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrPro 383
DB 115991 -----ACGGTGTCTGACAGCATTTTTCACAGACACGACGATGCTAGCC 116035
QY 384 GlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSer 403
DB 116036 GCCGCTGAATTTACTGGCTTCAGATCAACAGTATGATGTTCTGTGTGAGGACTTATTCC 116095
QY 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
DB 116096 AACGACGACACCCCAATTTGCTCGCCAAATCTGATGTTTAACTTCTTAAACCCAGAGGG 116155
QY 424 LeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
DB 116156 CTGTCCTCCGAAGGGGATAACGTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 116215
QY 444 GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463
DB 116216 GGGAGTCCGGTAGCGGTGGCTTCGGCAGCTTACCAAGTGGTGGCTGGAATCCTCTAAC 116275
QY 464 ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483
DB 116276 GTGATTTGATAAGAGTTGGTGAACATGATTTGTGGCAACGTAACCTACTATCAGTCCAC 116335
QY 484 SerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
DB 116336 GCGCAGACCATCAAAACCCAGATCAGATCTTACAACTCTGGTTAGCTCGCT 116389

RESULT 12
AB011450/c 10029 bp DNA linear BCT 28-APR-2003
LOCUS Leptosira interrogans serovar lai str. 56601 chromosome I, section
DEFINITION 259 of 397 of the complete sequence.
ACCESSION AB011450 AB010300
VERSION AB011450.1 GI:24196692
KEYWORDS Leptosira interrogans serovar Lai str. 56601
SOURCE Leptosira interrogans serovar Lai str. 56601
ORGANISM Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE 1 (bases 1 to 10029)
AUTHORS Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., Zhang, Y. X.,
Xiong, H., Lu, G., Lu, L. F., Jiang, H. O., Jia, J., Tu, Y. F., Jiang, J. X.,
Gu, W. Y., Zhang, Y. Q., Cai, Z., Sheng, H. H., Yin, H. F., Zhang, Y.,
Zhu, G. F., Wan, M., Huang, H. L., Qian, Z., Wang, S. Y., Ma, W., Yao, Z. J.,
Shen, Y., Qiang, B. Q., Xia, Q. C., Guo, X. K., Danchin, A., Saint

```

Girons, I., Somerville, R.L., Wen, Y.M., Shi, M.H., Chen, Z., Xu, J.G. and Zhao, G.P.
 Unique physiological and pathogenic features of *Leptospira* interogens revealed by whole-genome sequencing
 NATURE 422 (6934), 888-893 (2003)
 12712204
 2 (bases 1 to 10029)
 Ren, S., Fu, G., Jiang, X., Zeng, R., Xiong, H., Lu, L., Lu, G., Jiang, H., Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y., Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and Zhao, G.
 Direct Submission
 Submitted (12-MAR-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China
 Updated information will be available at our World Wide Web site (<http://www.chgc.sh.cn/lep/>). Comments to the authors are appreciated.

Location/Qualifiers
 1..10029
 /organism="Leptospira interrogans serovar Lai str. 56601"
 /mol_type="genomic DNA"
 /strain="56601"
 /serovar="Lai"
 /db_xref="taxon:189518"
 /chromosome="I"
 /note="Strain 56601 is maintained by the Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine"
 complement(148..1542)
 /gene="flgE"
 /locus_tag="LA2848"
 complement(148..1542)
 /gene="flgE"
 /locus_tag="LA2848"
 /codon_start=1
 /transl_table=11
 /product="flagellar hook protein"
 /protein_id="AA050047.1"
 /db_xref="GI:24196693"
 /translation="MRSISGVSLGNHQMVDVIGNNISNVNTHGKTERVTFQDM
 ISQELRGAPKENTGVNYPQVGLSLIAADKIMTQGLTCKTKNDVAISGEGFP
 IVKDKQVPTAGFLDKNGVYVNPANGLVQVNSLDDKNGKYNRSASIEDII
 IPVYSEKPARASQIDFKSNLSAPAPPDATQBEITAMINDPPKMRGHTVINT
 PDQGIQRFEKFEIKVRDNTWAKLSMTDQLSDVSGTGQDQLPQNTSEIFGF
 TPDKLVYSDGVSDMNGKLNKAKVSRIPGNPAIONFDLNLGEAGVNGITQFSSDF
 TTKAVQDGYTGYLESIDNSGITGVFSGVGRQPLARVATAVFNPNAGLDKAGDT
 MFSYNSNGEPNIGAGVQGRGKINAGLLENSVLDLSDQFTDMIVTQRFQAGRTIT
 TSDQMTQEVILGLKR"
 complement(1570..2364)
 /gene="flgP"
 /locus_tag="LA2849"
 complement(1570..2364)
 /gene="flgP"
 /locus_tag="LA2849"
 /codon_start=1
 /transl_table=11
 /product="flagellar hook assembly protein"
 /protein_id="AA050048.1"
 /db_xref="GI:24196694"
 /translation="MPEATGISQANINRYLEGDRSFNIRNMHENLEKEENGLKGI
 IRSTVSLGKDDPLKLLITQLSSQDTPNVKQDFIAQMAQFSSLEOMNNISLQKQ
 GNRQSFSLGKLVSPDFVNGENIAGIAGALPFDCEGKTFYRVNGRSIDVQISLSD
 PVVLKEQEAAYNQSSQSPARTPAHSEIGTVSPQEAPEIDISSNAHEMTTKQKSM
 ESSINKQREVFGHEQPELQEKITSEKISDWDKFPFGKDSNSTE"
 complement(2379..3944)
 /locus_tag="LA2850"
 complement(2379..3944)
 /locus_tag="LA2850"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA050049.1"
 /db_xref="GI:24196695"

translation="MNTSGDLSYSEKPPQSMESLPMGEVAGKNSFVLMKFLQSGA
 QKGLTEALQNTSSKAELQPDEEDPEIIEKKKRSINSIEEISKEENFTTQBA
 NNANILPFWLSETKTNESVDPEIESVILDEQSEIKTKNSTETLNAPEISSEILIE
 NFANEVEESTKLLSEVLSKEELNLDTEKSLVSPENLKEIKARSTKQESIYEKES
 ESFEAKDSOTFRFNEVDKKEKENTKGLSNFCKDFEAKVEKVELALDSEKMKKI
 SRDKTSDYNTAKAEEIKAVVLNQFTENSFKGSGOESPRSDSYSLVKGVGTTA
 ANGRESGTFSKEFLPSKESNVLSSKDDIQONFONLIRSAVQILENGRTEASIRMNPK
 DLQMSLAISTDKDVRGKLLVSDTVKQQLVAELANLKQDLKANGLELESILVIEFK
 REGFSFNADSKNGKDSHSAFQAFGDMNSDFKHSFYEEDELSEFENFSEPNDFPKK
 TDKSEKLLDKV"
 4359..4496
 /locus_tag="LA2851"
 4359..4496
 /locus_tag="LA2851"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050050.1"
 /db_xref="GI:24196696"
 /translation="MLSPFYFWRGGGKVKRFLYQKNILFANKKDSFLSEHLKNVC
 P"
 complement(5417..5797)
 /locus_tag="LA2852"
 complement(5417..5797)
 /locus_tag="LA2852"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA050051.1"
 /db_xref="GI:24196697"
 /translation="MDVSYFFGYATTQYVSIPFILKEYSVRKPELENGDFQIBAEFPS
 YLHTCKKQSYRDFDSGLIRINDYTADIVGSNAKGAHFTTHFDQIEGFPNATKRDVEV
 RLGSWATSPVLGAELKPLGVYFR"
 complement(5898..6128)
 /locus_tag="LA2853"
 complement(5898..6128)
 /locus_tag="LA2853"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050052.1"
 /db_xref="GI:24196698"
 /translation="MDPFFXIVNRYGGWELWNLDFVEFYFQSLSGPLFLKGLGKTF
 NKGKGVFPKFEITFDLPFNPKNNIQKW"
 complement(6485..7084)
 /locus_tag="LA2854"
 complement(6485..7084)
 /locus_tag="LA2854"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050053.1"
 /db_xref="GI:24196699"
 /translation="MRKIMILLPLCLVFASCEDEKKDGSITGNSIMDLLLQEVST
 PPGPGVWMDIPSTINAQVTSKPSFIQFSAGSVNHTLMKNKCNFSELSVTNL
 PAGLTLNSTGAINGAPTASATVTTSKAKKANNSPTITFTKTTVTIVFAAGSLTC
 NTAGALGCNNALPYSCNSNFCYSTYSSCKAASECGY"
 7397..7531
 /locus_tag="LA2855"
 7397..7531
 /locus_tag="LA2855"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050054.1"
 /db_xref="GI:24196700"
 /translation="MIFINKITPKKLTFCERIFRIHFKMKNLKASLILFTKYRKTKQ
 K"
 complement(7860..8000)
 /locus_tag="LA2856"
 complement(7860..8000)
 /locus_tag="LA2856"
 /codon_start=1

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AA050055.1"
/db_xref="GI:24196701"
/translation="MTLLNDIYCKRIKIEAKINEQCKDVEILENWSRNSRFPRIIL
SF"

```

```

gene      8110.. 9156
          /locus_tag="LA2857"
CDS       8110.. 9156
          /locus_tag="LA2857"
          /codon_start=1
          /transl_table=11
          /product="predicted hydrolases or acyltransferases,
          alpha/beta hydrolase superfamily"
          /protein_id="AA050056.1"
          /db_xref="GI:24196702"
          /translation="MSCLSKSYBEKRMKILKWLGIITGTFALFVVTYATPKVEY
          KSVPLYSNFDYRKLQYSRKKVPGNEELKLYSADKTDPSILYIHGPGASRAEG
          EYTDOLAKDKANLYKPLGHTNLEHRDITFEILQDSETPLECEKLGKTKIL
          IGTSMGLISTYLAAYKPKVHALVLPSPDYFTNPFVSIYOFQSWGKDFANIVMGKIR
          KSTEEKRNPAFAFYRDVYLAAVONLSDLREIFLCTDPFSKISPIILLFYIKNEKN
          QDVSAVSVMNLAKVKNENGKASPLNKAIVKVEFGNHLVFSKYMKSDDKLLIKETFP
          IQNVFSLNKLNSH"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      9,05e-26      Length:      10029
Score:          550.00        Matches:      149
Percent Similarity: 43.63%    Conservative: 84
Best Local Similarity: 27.90% Mismatches:      199
Query Match:     21.27%      Indels:       102
DB:              12          Gaps:         12

```

US-10-009-823A-1 (1-502) x AE011450 (1-10029)

```

Qy      1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Dy      1542 ATGATGAGGTCACCTCTATTCGGGTGATCCGGACTTAAGAACCCACCGTCCGAATGGAT 1483
Qy      21 ThrValSerAsnLeuAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Dy      1482 GTCATTGGAAACACATCTCCACGTAACACTCAGCGTTTAAACCCGACGGGTACG 1423
Qy      41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Dy      1422 TTTTCAGGATATGATTTCTCAGGAGCTTAGAGGTGCTCCGAGCCTAAAGAGAAATATTGGA 1363
Qy      56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Dy      1362 GGTGTGAATCCTCACAAGTTGTTGGGTCTTTGATCGCTCGGATTGATAAATCATG 1303
Qy      76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
Dy      1302 ACTCAAGGTTCCCTTCAACACCACTGGTAAATAATACGACGTTCCGATTTTCAGCGCAAGGT 1243
Qy      96 PhePheGlnValThrLeuGluAsnLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Dy      1242 TTTTTCATCGTCAAGGACCGGTGACAAACAATTTTATACGAGACCGCGTCTTTTAACTTC 1183
Qy      116 ThrGlnAspGly----PheLeuAsnAspProSerGlyPheThrLeuMetGly-----Ser 132
Dy      1182 GATAAAACCGTTACTACGTAACCCCTGCTAATGTTCTTAAGGTTCAAGGTTTGGAACTCT 1123
Qy      133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152
Dy      1122 AGACTGGATGATAAGGAATAATAATCATCAATTCGGACGCTTCGATTGAA---GATATT 1066
Qy      153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
Dy      1065 ATAATTCAGTATATTTCTAAGAACCTGCAAGGCTACTTCTCAGATAGATTTTAAATCC 1006
Qy      173 AsnLeuGlyAspSerThr-----AspLysThrGlnSerGluAlaAsnPro 187

```

```

Db      1005 AATTGGAATCTCTGCTCCGGCGGTTCCGCTGACGCAACCAAGAAAGAAATCACT--- 949
Qy      188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207
Dy      948 -----CGCATGATTAAACGAT-----CCAGATCCTAAGATGAGA 916
Qy      208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlnGlyAsnSerHisAspIleThr 227
Dy      915 AGGGACATGTTACCACGATTAAATACATTTCAGCATCAAGGATTCAAGAGAAATTTAAA 856
Qy      228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Dy      855 ATGGAATTTTCAAGGTTTCGTGATAACACTTGGAGGCTAGATTGACGTATGACGGATTCA 796
Qy      248 MetAsnProSerGluAspGlySerAlaAspSerGlyThrAspSerAla----- 263
Dy      795 ACTCAACTTCCGTTGATGTTCTGGACCGCGGCAAAACACTCACTCACTCCAGGTAAC 736
Qy      264 -----GlyLeuLeuMetSerGlyThrMetThrPheSerSer----- 275
Dy      735 ACTGAATTAGAAATTTGCTTTTACTCCGGATCGAAACTAGTATACGTTTCTGACGGTGTG 676
Qy      276 -----AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer--- 290
Dy      675 GATTCGATGAACACGCGTAACTTGAACGCAAAAGTTTCTTTTCGGATCTCTGGAATCCT 616
Qy      291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuPro 308
Dy      615 CCGATTCAAAATTTGATCTTAACCTTAGGAGAA---GCAGGAATGGTAAACGGAATCACT 559
Qy      309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle 328
Dy      558 CAATTCTCCTCTGACTTT----- 541
Qy      329 LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Dy      541 ----- 541
Qy      349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Dy      541 ----- 541
Qy      369 AsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuVal 388
Dy      540 -----ACTCAAAAGCTGTCAACACAGACCGTTATACGATGGTTCATCAGAA 493
Qy      389 AspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
Dy      492 TCTTTTCGATCGATAATCTCGAACCGATTACCGGTGTGTTTCGACCGGAGTCGCTCAG 433
Qy      409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGly 428
Dy      432 CCTCTTCGATGAGTTGGACGCTCGGTTTAAATATCCCGCGGTTTAGATAAAGCGGGT 373
Qy      429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
Dy      372 GATCAATGTTTTCTGATTCCATCACTCGGAGAACCTAACATCGGTGAGCGGAGTC 313
Qy      449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Dy      312 CAAGAAAGAGTAAATCAACGACGAGGACTTTAGAGATGTCATATGTTGATTCCTCCGAT 253
Qy      469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Dy      252 CAGTTTACGGATATGATCGTGACTCAGAGAGGTTTTCAAGCTAACTCTAGAACGATTACT 193
Qy      489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Dy      192 ACTTCCGATCAGATGATTCAGAGAGGTTTGGGACTCAACGCT 151

```

RESULT 13

AE0117291

LOCUS

AE0117291 300594 bp DNA linear BCT 29-MAR-2004

[illegible]

```

/db xref="GI:45600125"
/translons="MNNRILESXYTYLENQKJFLSYSETDSKRVILFVHGYPDTH
KTWDLQVNLKERTGAIIDRIGSGRSKSPSESEYNTMIISDLLEAIRFLSKKKV
HLVCHDGAALGWLFSDPFSYKVSFTAISGPHPLAGKRIIDDLFSFSLSNWSKV
IDGFSWYIWFQIPVLPPLIWNFGESYKWMIDGVPKDSLRKKNKNDIYNST
IAPINLRELLFKTYVLISSPSHKKVPQVLPVQKDFIVLPEVYENSIDYVEKLEIHKI
PSNVHREQPDVLELIDQKVFVKIS"

```

gene

```

/locus tag="LIC11001"

```

```

/notes="transposase; IS1533; this region contains an
authentic frameshift, and is not the result of a
sequencing artifact"

```

gene

```

/locus tag="LIC11002"

```

```

/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

```

CDS

```

/locus tag="LIC11003"

```

```

/product="transposase, IS1533"
/protein_id="AA069612.1"
/db xref="GI:45600126"
/translons="MNNLHPENILQETFDNYYSRVROEENLKAMDQKIOIEASEP
FKERVGLRCFRGVDTLT"

```

gene

```

/locus tag="LIC11003"

```

```

/complement(9714..11381)
/gene="lipL71"
/locus tag="LIC11003"
/complement(9714..11381)
/gene="lipL71"

```

CDS

```

/locus tag="LIC11004"

```

```

/product="lipL71"
/protein_id="AA069613.1"
/db xref="GI:45600127"
/translons="MKSOTKSLPFIILLAGFFISGALPIEELSDAKNSITRAKS
AGAEKAPSLERAKNLTATOHKASEENITETKSAVARAKALDASEKSPSPVDD
ARKESSSIESAEVASOLASFPYNTSVOLKEGSLBETADRTLESYKESGDGAK
LRMLAFDQYIEARQKIDAKADESVLALSQKQLIDSPADIKNLNADKVAE
GKQPEVSETRLDSSKLEEGKEGYSEIDIRKSGELVAKNIKILYAEQKELA
KQVSAATTLASFDNRKNSRDFQVSYORAEENLKAAEERVAEDLYSEKVEDS
ISRSEAIRLSRILVDQATLARIERKATTDIAGRTKTENKNTKNQSTTEGNS
SXIIGEDLPKGRVYVKKVPADCLMRIAKDRHYGTSLWRITYEARNRKIKNPN
LIVPKOVLILPRKGTPLDKVESAPSRKKPATEVEALEENRKPETHSESDSEAGE
SNKKKABATPPTDSEERGSESENEBSEENGEENPENLQ"

```

gene

```

/complement(11429..11764)

```

CDS

```

/locus tag="LIC11004"

```

```

/complement(11429..11764)
/locus tag="LIC11004"

```

```

/notes="identified by sequence similarity; putative; ORF

```

Alignment Scores:

```

Pred. No.: 5,09e-24 Length: 300594
Score: 550.00 Matches: 149
Percent Similarity: 43.63% Conservative: 84
Best Local Similarity: 27.90% Mismatches: 199
Query Match: 21.27% Indels: 102
DB: 12 Gaps: 12

```

US-10-009-823A-1 (1-502) x AE017291 (1-300594)

```

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetIysThrHisSerThrGlyLeuGly 20
Dd 251203 ATGATGAGGTCACCTATTCGGTGTATCCGACTTAAGAACCAACCAAGTGGAT 251262
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Dd 251263 GTCAATCGGAACAACATCTCCAGCTTAACACTACCGGTTTAAACCGGAGTACG 251322
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55

```

```

Dd 251323 TTTCCAGGATATGATTCTCCAGAGCTCAGAGCGCTCCAGAGCTTAAGAGAAATATTGGA 251382
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Dd 251393 GGTGTAATCTCAACAAGTTGGTTCTTTGATCGCTGCGATTGATAAATCATG 251442
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyValGly 95
Dd 251443 ACTCAAGTTCCCTTCAACACCACTGGTAAATACGAGCTTGGATTTCAGGCAAGGT 251502
Qy 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Dd 251503 TTTTATCGTCAGGACGGTGACAAACAATTTATACGAGAGCGGTCTTTTAACTC 251562
Qy 116 ThrGlnAspGly----PheLeuAsnAspProSerGlyPheThrLeuMetGly-----Ser 132
Dd 251563 GATAAAACCGGTTACTAGTAAACCTGCTAATAGTCTTAAGGTTCAAGGTGGAACTCT 251622
Qy 133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152
Dd 251623 AGATCGGATGATAAGGAATAAATACATCAATTCGGCAGCTTCGATTGAA---GATATT 251679
Qy 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
Dd 251680 ATAAATCCAGTATATTTCAAGAACCTGCAAGGCTACTTTTCAGATAGATTTTAAATCC 251739
Qy 173 AsnLeuGlyAspSerThr-----AspLysThrGlnSerGluAlaAsnPro 187
Dd 251740 AATTGAAATCTTCTGCTCGCGGTTCCGCTGACCAACCAAGAGAAATCACT--- 251796
Qy 188 TyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProIleSerThrSerAsn 207
Dd 251797 -----CGGATGATTAAAGAT-----CCAGATCTCAAGATGAGA 251829
Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227
Dd 251830 AGGGGACATGTTACACGATTAATATACATTGACGATCAAGGATCAAGAGAAATTTAA 251889
Qy 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Dd 251890 ATGGAATTTACAAGGTTCTGATAACACATTGGAAGCTAGATTGAGTATGACGATTCA 251949
Qy 248 MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla----- 263
Dd 251950 ACTCAACTTTCCGTTGATTTCTGGAACCGCGGCAAAACACTCAACTTCCAGGTAAC 252009
Qy 264 -----GlyLeuLeuMetSerGlyThrMetThrPheSerSer----- 275
Dd 252010 ACTGAATTAGAAATTTGGTTTACTCCGGATGGAATACTAGTATACGTTCTGACGGGTG 252069
Qy 276 -----AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer--- 290
Dd 252070 GATTTCGATGAACAGCGGTAGTTTGAACGCAAAAGTTCTTTTCGGATTCTCTGGAATCT 252129
Qy 291 -----AlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuPro 308
Dd 252130 CGGATTCAAATTTTGTATCTTAATCTTAGGAA---GCAGGAATGTTAAACGGAATCACT 252186
Qy 309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle 328
Dd 252187 CAATTCTCTCTGACTTT----- 252204
Qy 329 LysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Dd 252204 ----- 252204
Qy 349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Dd 252204 ----- 252204
Qy 369 AsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuVal 388

```



```

/ gene="fliM"
3865..4899
/ gene="fliM"
/ note="flagellar switch protein"
/ codon_start=1
/ transl_table=11
/ protein_id="AAB61256.1"
/ db_xref="GI:2196898"
/ translation="MTSEVLSDQDIDQLLTALSSDASIEDARPISDTRKITLYDRP
DKPSKQMRILSLMHEFALTTLSLAQRLSMHVHVASVDLTVEERISPTPTST
LAVITMPLKGNVLEVPDPTTISIDRLFGGQAAKQVORDLIDENSVEGVIRI
LANVRESWTQIDRLPELGGIETNPQFAQIVPPEMVLVLETKVGEEMGMFCIP
YITTEPTISKLSSQFSPSVRSSTQTMGLVRLKLSLTDMDVVAEVSRLSLVRDIL
GLRVDTIIRLHDTHVGDPPFVLSIGNRKFLCPQGVVKKIAQILERIBSTSEDPEE
LSADEEELYE"
4880..5230
/ gene="fliY"
/ gene="fliY"
4880..4883
/ gene="fliY"
4892..5230
/ gene="fliY"
/ codon_start=1
/ transl_table=11
/ product="fliY"
/ protein_id="AAB61257.1"
/ db_xref="GI:2196899"
/ translation="MSDGSISQABIDALLSGVEMGADSPAGASFGGSLSAQSEL
LQSPQANVGSPFSLSLGKTQVTSIPPRVEISSKEFSRVIPDMLNATVIDFEGEP
AGDHLFLWAP"
ORIGIN
Alignment Scores:
Pred. No.: 4,25e-25 Length: 5230
Score: 534.50 Matches: 154
Percent Similarity: 42.70% Conservatives: 80
Best Local Similarity: 28.10% Mismatches: 183
Query Match: 20.67% Indels: 131
DB: 1 Gaps: 14
US-10-009-823A-1 (1-502) x TPU28219 (1-5230)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 22 ATGATCGGTCACCTTTTTCAGGTGTCTCTGGTATGCAAGATCATCAACGCCCATGGAT 81
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 82 GTCAATCGGAACACGTCGGGACGTTTAACTACCGGTTTAAAGCGTGGCGGTGTTAAT 141
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
DB 142 TTTCAAGATCTTATTTCTCAAGCACTGAGTCGGCTCGCGGTCCGATGAGAGTTGA 201
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
DB 202 GGAGTGAATCCCAAGGAAGTGGGATTTGGCGTGTCTGATTCAGACGATCGATCTGTTAC 261
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
DB 262 ACGCAGGTGCACTGCAACACGCGGTATCAATACGATGTGTCTATTACGGGAGTGT 321
QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
DB 322 TTTTGTGTGAAAGTGGGAAAAGACCGTTTTCACCGCGCAGTGCCTTTGGGTT 381
QY 116 ThrClnAspGlyPheLeuAsnAspProSer-----Gly 126
DB 382 GATAATCGGCGACTCTCTGGAACCTTCGGAATGGTATGCGCGCTTCAAGTTGGATGGCG 441
QY 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlnThrLeuGlu 146
DB 442 CAGGACGTGGCGGGAGCGGTTTAAATTAATTCCTCTGCACAGACGAGGATCTCGTTATC 501
147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
DB 502 CCCATTGGG-----CAAAAGATAGATGCCGACAGCACCGACG 537
QY 167 AlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
DB 538 ACTGTCTACTATGCCCTGTAATTAGACAAGCGCTCTGCCCTGAGCTTGTCTGAGATGCCAAC 597
QY 187 ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206
DB 598 GAACCGGACGTCGTAAGTCCACGTGG-----ACAACT 630
QY 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
DB 631 GACTTT-----CAAGTGATATGATAGCTTCGGGACGACGATACGTTG 672
QY 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
DB 673 CAGATTAACTTTTCCGCTGTCGCGGACG-----AACAAATCAGTGGCAGCCCATGTC 726
QY 247 AlaMetAsnPro-----SerGluAsp 253
DB 727 GCAGTGGATCCGGGACAGAGTAGATACGCAACGCGGTGAGGGTGGGACATCTGAC 786
QY 254 GlySerAla-----AlaSerGlyThr 260
DB 787 GGTGGCGGCAACACCTTTTATTGTAATTTTGGACACCTCGCTTCAGTGACT 846
QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
DB 847 GACACTGGACGG-----858
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
DB 859 AACGTGACCGGT-----CCTACCGGACAGGTG-----885
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
DB 886 -----CTCCTTGAAGCGTCTGACGATGTTGTCGTGG-----918
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
DB 919 -----AATCCGACGATGACGAGCGAGT 942
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
DB 943 ACGCGCCACGCTTTCACGCTCACTTGGTGAAAT-----978
QY 361 SerSerGlyAsnSerThrAlaArgAsn-----GlySerSerSerThr 374
DB 979 -----GGCACCGCGCGCAATACGATTACGCGATTTGCTGAACGCGATCTACC 1026
QY 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
DB 1027 AAACCTTACCGGACGAGCGGTACGCGATGGATATTGAAATTTTAAATATAGATCAA 1086
QY 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
DB 1087 AGCGGTGTCATCTGCTGCTGTTTCAATGGGTGAGCCAGACATTTGCCAGCTCGCA 1146
QY 415 LeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAla 434
DB 1147 CTTGAGGATTGCAATCAAGGTGGTCTTGAGAAGGACGAGAGAACACCTACGTACAA 1206
QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
DB 1207 TCGAACAACTCAGGATAGCGAACATTAGCACCTCGGGGGTGGAGGAGGAAAGTTG 1266
QY 455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
DB 1267 ATTGAGGAGCACCTTGAGATGACGACGAGTATTAACCGATCAATTTACGATATGATC 1326
QY 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetLeu 494

```

```
|||||
1327 ATTACCAAAAGGGTTTTCAGCGGCGCAAGATTCAGACATCAGACACCATGTTG 1386
|||||
495 GlnLysAlaLeuGluLeuLysArg 502
|||||
1387 GATACCGTGTGATTTGAAGCGC 1410
|||||

RESULT 15
AB001244/c
LOCUS
DEFINITION
Treponema pallidum subsp. pallidum str. Nichols BCT 10-OCT-2003
the complete genome.
ACCESSION
AB001244 AB000520
VERSION
AB001244.1 GI:3323005
KEYWORDS
SOURCE
Treponema pallidum subsp. pallidum str. Nichols
ORGANISM
Bacteria: Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE
1 (bases 1 to 26536)
Praser, C.M., Norris, S.J., Weinstein, G.M., White, O., Sutton, G.G.,
Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,
Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,
Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M.D.,
Fujii, C., Garland, S., Hatch, B., Hurst, K., Roberts, K., Sandusky, M.,
Weidman, J., Smith, H.O. and Venter, J.C.
Complete genome sequence of Treponema pallidum, the syphilis
Spirochete
JOURNAL
Science 281 (5375), 375-388 (1998)
MEDLINE
98332770
PUBMED
965876
REFERENCE
2 (bases 1 to 26536)
Praser, C.M., Norris, S.J., Weinstein, G.M., White, O., Sutton, G.G.,
Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,
Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,
Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M.D.,
Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Watthey, L.,
Weidman, J., Smith, H.O. and Venter, J.C.
Direct Submission
Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 26536
Location/Qualifiers
/organism="Treponema pallidum subsp. pallidum str.
Nichols"
/mol_type="genomic DNA"
/strain="Nichols"
/sub_species="pallidum"
/db_xref="taxon:243276"
complement(100..303)
/gene="TP0701"
complement(100..303)
/gene="TP0701"
/notes="similar to GB:U10328 SP:P08374 GB:M15266 GB:M24503
PID:146978 percent identity: 27.59; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase, putative"
/protein_id="AAC65670.1"
/db_xref="GI:3323006"
/translation="MIFPMQQLIEFGNIYEITCAATRAAQLAAVCDPVLDELGGKV
VSAQAQQVSGTVDYRIEPELUG"
383..456
/gene="trRNA-Pro-3"
383..456
/gene="trRNA-Pro-3"
/product="trRNA-Pro-3"
/notes="codon recognized: CCC"
complement(473..1009)
/gene="TP0702"
complement(473..1009)
CDS
/gene="TP0702"
/notes="similar to GB:AE000783 percent identity: 31.72;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC65671.1"
/db_xref="GI:3323007"
/translation="MNSPQTRAPLGPQRQVESRRCTSVFGQARVYVYVDTGTSQAVH
WGDGVPVAVTAAGTGRVLAEGRTTGWSVLEHAGPLGYTAHNLHLLKKDT
YVSTGTLIASGTTGTFSTGPHVWEARINSTPIDPECLLAPPLTTLHPKVALRKLOPT
GETQGPPLPKDNLGPG"
complement(1024..1704)
/gene="TP0703"
complement(1024..1704)
/gene="TP0703"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0703"
/protein_id="AAC65693.1"
/db_xref="GI:3323023"
/translation="MGARVGRAPGVNAKRAVQTQVQSVPKSVKRVWVGLCEVLLF
SGVFSVAVSFLLSAAVPPPRQVPAHFRIAMQFPQVQSGACAVATSLVALQK
DKATVHLVLRTPAGKEAKTVRAFALPKGIAGLGTDAATAGVSRAGSQDVTHVALL
GISIFWEPGDWMLAEQVRVPGKPYVRRAPLRIEKKEFPRELELDKRNATIAQDKSE
RKVQRDA"
complement(1704..3824)
/gene="TP0704"
complement(1704..3824)
/gene="TP0704"
/notes="similar to GB:AE000783 percent identity: 24.42;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="single-stranded-DNA-specific exonuclease (recJ)"
/protein_id="AAC65672.1"
/db_xref="GI:3323008"
/translation="MKRWRAKALDEGRVRELTARYSCSALEAAILVRKVVADALLF
HFERDLALPGPFHFAHSAVDRLRAHERKEKVLIGDRDADGIGATLLFALCA
FGLTVKVRVPADPGLSCNADVEHAAAGTLITVDCGISNRAETAYARHRIQDVL
VIDHAAPELTPDAIINAKVPGAGYHEFLSGCATAKVVTALRFTSTGWYQRPVVL
LHVPVTGATIEAVKCNMEVDRLETSVVPBMLPSHTLGLAFSGQSVFVWDEOV
VRTQAKVGGEPFRTIDVRPMAVFPQACERASLLRMKDFSRIGARSKPMEIES
FINLITFVQKDTQLYEAGADELQVAISLSDMLVLAGENRINRVYRGLQMNAPRV
GLRELFATQRMGKGLTVEVGSIVPLINATGRIGCPHRAVELFLMREASQALQAR
KMWOLNEQRKLGSRARALVEPLARASLETYSNRLAVVCSDKTHRGVTGILANCLSEA
LRVPCVLI CIMADGHVGSLSARGVHLFSLIDPLADLFSDYGGHFAAGFSPISERI
PQLHRMELYAATIEFADESAQCEGFEADLDTQMTRGLLTLVDRPEPYGEGNPPLC
FLAKRLKIFASLFGRTERVHVKLTLDAQVHKPFIYWGAGEKLAQFAVGVDVDAVF
QVTRTFQGTCTPOLVIQDLWRSE"
complement(3967..6621)
/gene="TP0705"
complement(3967..6621)
/gene="TP0705"
/notes="similar to GB:AE000783 percent identity: 40.76;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein (pbp-2)"
/protein_id="AAC65673.1"
/db_xref="GI:3323009"
/translation="MRSVCVLYLVLLCGLLVGGAALFGLLASTENIKQSERFTRFNP
ALPRLIDIRGDLITESSDEKREIVSFADLSHLVHALLITREDSRYVTHGYSIKAI
FRALVGLTTRGALGGSTLTQIAGLLYSRDSRSLRRKIKELMWHLMHRRYSKNIH
NELYNRVYFGGTGVGAARAFYGHVSQMSAAEAALLILSNFAHNPFPYPNR
AQDRQVYVHMTRELGLYSEORDESIEYHNAHFDYTRTSSAFYABADKARVSEYV
RROLDRMYGTMNLYQDGYVHTCDLRLHQLVAEQVVEQLAQKASLTHVAR
RELYSNLSLGLVFNVPOLHVGDRMKAKSGAYRSTLGPLVNMALMFLDMLKTL
SDKASLVRDEIARKVEGALIALENDGYTLVAGGSRFGASQVIRATCGLLQPGS
VFKPLVYSAALDSKCLTNATQLHDAPQFNRNGVSIYENNYGKQWGLAWKALAS
LNIPIAIRVLDWMVGFDAVIQRAATLLHITDQIEIRTFPRVYPLALGVVALRPIQLARA
```

PAFNGCGKAVEPIAVSRVEDRLGRVILDPEREVRARLRQAQGAATOLISAENALMTN
 MLEKTVMTGLAVASERGRAFTYODPATGRSFVMPVAGKTGTTTONSDAWAVGYSYY
 TAVLWFGDGLSGHSTGATLGGPWFARFRAHQDHPRAFRVPTGLMYQAVC
 AKSLGTPYCGDELVSFLPGTVPFRFCSHESGKMLBELAKERLKSFGYSTGFWP
 AVSEGLHDPVRPEPDPGSPGVQGEVLVPSNPFAPVHPQEAFLDPAPVFPVSASD
 PSRDAPHESEDVQDDPARGVQSLDTGP"
 complement (6918..7844)
 /gene="TP0706"
 complement (6918..7844)
 /gene="TP0706"
 /notes="similar to GB:AE000783 percent identity: 36.30;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65674.1"
 /db_xref="GI:3323010"
 /translation="MVRAAVCSLCVCAHPVCGPYPLIAQLQGQDALFRQVSDVVR
 AARVALAQGRGNDPLRFYAVYKXADTIIRIAARCGIPYDAVASINRIETIKRELV
 GRTLLPTVGLVSDADPHLPRLRLIALIKKGSPFLSLPGTAQTTHTRKREVV
 CAPALPDGTVPRAFFLPKPYRFPLASGLRTSGFGARKSPETGRLSVHPCGIDLAAPGA
 LVYACAGQVATAYRNLKYLQHTDGRHSLYGHLNAYRVVQOKLSVGAVIGNV
 GSTAGTGPLHFEVRAGVQNPFRFMEXFR"
 complement (7906..8385)
 /gene="TP0707"
 complement (7906..8385)
 /gene="TP0707"
 /notes="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0707"
 /protein_id="AAC65694.1"
 /db_xref="GI:3323030"
 /translation="MRGTAYFAVSGVCSACTCTOVAVQLASWFGSLKRCREVR
 RRPALAIYVRVIRYAGLAMLAVLSFALGICALVYFLMALVHRFRVFSVLGLLYGG
 GAVLWGLRVCRNALSARVRARRAAQEPCLVQAVGEVANGULSSAEVRAPEEC"
 8451..8966
 /gene="TP0708"
 8451..8966
 /gene="TP0708"
 /notes="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0708"
 /protein_id="AAC65695.1"
 /db_xref="GI:3323031"
 /translation="MDVQERRFSCAAVGLKVPATAAGAAFFLSIATAAARNARVY
 VSVARATVLAGAGTAYALRALLAYVVDLLVQEDGKMPHPIHVDLTLDVVEPFAS
 PGDVQVETDSDSFTGIPTGELGELGYSFSPSPSPFEDKTSIDLVDLVDLPETKMA
 RAIRTVLSQDT"
 9022..9810
 /gene="TP0709"
 9022..9810
 /gene="TP0709"
 /notes="similar to GB:J03169 SP:PI7211 PID:153486
 PID:153487 percent identity: 47.79; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase sigma-28 factor"
 /protein_id="AAC65675.1"

Alignment Scores:

Pred. No.:	2,91e-24	Length:	26536
Score:	534.50	Matches:	154
Percent Similarity:	42.70%	Conservative:	80
Best Local Similarity:	28.10%	Mismatches:	183
Query Match:	20.67%	Indels:	131
DB:	1	Gaps:	14

US-10-009-823A-1 (1-502) x AE001244 (1-26536)

1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 26471 ATGATGCGCGTCACCTTTTTCAGGTGTGCTGATGCAAGATCATCAACCCGCGAT 26412
 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 26411 GTCATGTGGAAACAGCTCGCAACGTTAACACTACCGGTTTAAAGCGTGGCGGTAAAT 26352
 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 26351 TTTCAAGATCTTATTCTCAGCAACTGAGTCGCGCTGCGCGTCCGAATGAAGAAGTTGGA 26292
 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 26291 GGAGTGAATCCCAAGGAAGTGGGATGGCGTGTGATGCAACGATGATCTGTTTTCAC 26232
 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
 26231 ACGCAAGGTGCTGCAACACGCGGTATCAATACGATGTGTCTATTTCAGGAGAGTGT 26172
 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 26171 TTTTGTGCTGAAAAGTGGGAAAGACGTTTTCACCGCGCAGGTGCTTTGGGGTT 26112
 116 ThrGlnAspGlyPheLeuAsnAspProSer-----Gly 126
 26111 GATAATGCGGCACCTCTCGTGAACCTCGCAATGGTATGCGCGTTCAAGTTGGATGGCG 26052
 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu 146
 26051 CAGCAGCTGGCGGGAGCGTGTAAATTAATTCCTCGCAGACGAGGATCTGTTATC 25992
 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
 25991 CCCATTGGG-----CAAAGATAGTAGTGGCGCAGACAGCAGC 25956
 167 AlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
 25955 ACTGTTCACTATGCTGTAAATTTAGACAAGCGTCTGCTGAGCTTGTCTGAGATGCGAAC 25896
 187 ProTyrPheAlaLeuLeuGluSerThrLysGlyAsnGlyThrProIleSerThrSer 206
 25895 GAAGCGGACGTGCTAAGTCCACGTG-----ACAACT 25863
 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
 25862 GACTTT-----CAAGTGTATGATAGCTTCGGCGCAGCAGCATACGTTG 25821
 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
 25820 CAGATTAACTTTTCGCTGTCGCGGGGACG-----AACAACTCAGTGGCAGCCATGTC 25767
 247 AlaMetAsnPro-----SerGluAsp 253
 25766 GCAGTGGATCCGGGACAGAGGTAGATACGAAACGCGTGTAGGGTGGGACATCTGAC 25707
 254 GlySerAla-----AlaSerGlyThr 260
 25706 GGTGCGGCAACACCTTTATTGATAAATTTTGGACACCTCGCTTCAGTGACT 25647
 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 25646 GACACTGCAGG----- 25635
 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrGlnPro 300
 25634 AACGTGACCGGT-----CCTACCGGACAGGTG----- 25608
 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 25607 -----CTCCTTGAAGCGTGTACGATGTTTGTGCGTGG----- 25575

QY	204	SerThrSerAsn-----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGln	220
Db	31592	AATGTAGCTAATCCAGATACCTATACCCATTCTACTTCGGGAGAAAATTTTGACAGATTTT	31651
QY	221	GlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThr---	239
Db	31652	GGTAATATCCCATPATATAATTTATTTTCAGAAAACAGCTGCATCCATCTGGGGCGGTA	31711
QY	240	LysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly	259
Db	31712	TATACAGCCGTGGATGGTACAGTAGTAGTACCTCT-----ACTGGATTATCTGGT	31762
QY	260	ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeu	279
Db	31763	ACGGATAGC-----CCAACTACTTTGAGCTTTAATCCATCCGGG-----	31801
QY	280	LysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGln	299
Db	31802	-----ATCATGACATCAACCGGTCAAGTACGATTATGTG-----	31837
QY	300	ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIle	319
Db	31838	-----GCTGCTTCTGTACTGGTGGCGGGTT	31864
QY	320	GlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaPro	339
Db	31865	TCACCGCTTGCTTTGAGCTTGATTAAAC-----	31894
QY	340	AlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGln	359
Db	31894	-----	31894
QY	360	ThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGln	379
Db	31895	-----CGTTCCACCAGTTTGGCAGCAAAATTCGGGGTGAATGTGATGAGCCAG	31942
QY	380	AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln	399
Db	31943	GATGATTGTCTTCGGGTGCGCTGGCCGGGTACAGCAATCTCTGATGTGGCCACATCAGA	32002
QY	400	GlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThr	419
Db	32003	GCACGTTACTCCAATAACATCAACGCGCACAGTCGCGTCAACTGTACTGGCAGGTTTTC	32062
QY	420	SerGluAspGlyLeuArgArgGluGlyIleAsnAsnHisTyrSerAlaThrLeuAspSerGly	439
Db	32063	AATCTTAACGGTCTCAGCCGCTGGAGAGCAATCAGTGGGGTGAGACGGCTGATCCGA	32122
QY	440	GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu	459
Db	32123	TTGCCCTTGTGGGGACTCCGAGACGGGTGCTGGGCTCACTGCATCTCTACGAGTG	32182
QY	460	GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGly	479
Db	32183	GAAGAATCCAATGTTGATCTGACGACCGAGCTGGTCACCCCTGATCATGACCACCGGGTC	32242
QY	480	PheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGlu	499
Db	32243	TATCAGGCCAATGCCAAGACCATCGAGACACAGGATGCAGTAATGCAGACCATCATGAAT	32302
QY	500	Leu 500	
Db	32303	ATC 32305	
RESULT 17			
AE013102/c			
LOCUS		10303 bp	DNA linear BCT 10-JUN-2004
DEFINITION			Thermoanaerobacter tengcongensis MB4, section 129 of 244 of the
ACCESSION			complete genome.
VERSION			AE013102 AE008691
KEYWORDS			AE013102.1 GI:20516444

Db 1091 GGACAAACCCCTCAGCAAAATGGCTGGGAGTTGCCATAGCTCTTATTGATCTTTATTT 1032
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 1031 ACAAGAGGAGGCCAGAGAGAGCTGACAAACCAAGATTTGTCATTGACGGAATGCA 972
Qy 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 971 TTTTTCATAGTTTCAATGGGAATCAATTTCTTTACAGGAGCGGGAATTTTATGTTT 912
Qy 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
Db 911 GACTCTAATGAGACATTTGGTAACCTCTGATGGATATAAAGTTTGTAGATGGATCACT 852
Qy 136 AsnAsnProAsnIleLysLysGluThr-----LeuGluProIleGlnLeuAsp----- 151
Db 851 GATGGCAAGACAGTACACAGTACACCGGGAATTTGGTTCCGATAAGCTTAAAAAAGTGG 792
Qy 152 PheAsn---AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 791 TTTGGGATGGAGCTTCT-----ACCACTACTCAATTAGAAATA 753
Qy 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190
Db 752 GCGGGAATTTGAATGCTTCTACTGGAATAGTGCAGAGC----- 714
Qy 191 LeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
Db 713 -----ATAAGCTAC 705
Qy 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
Db 704 -----AACATAATCGTCTATGACACACAAAGGTGGAAGCCATGTGCTACAATACTTTC 651
Qy 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
Db 650 ACAAGCAAGACCCCAATACCTGGAAC-----TGGAGTGTA 615
Qy 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr 270
Db 614 TCTTCTCAAGACCGGTATATTTCTTCTGCTGCTGCTGCT-----TCAACGGCA 564
Qy 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
Db 563 ATCACTTTTGGAGCTGATGGAAGATAGTCTCTAATACG-----CAGGCTACCGGAATC 510
Qy 291 AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe 310
Db 509 TTAACCTTTAATCAATGAACA----- 489
Qy 311 SerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSer 330
Db 489 ----- 489
Qy 331 GlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGly 350
Db 488 -----CGAGTTCAAAATGCTCAGATTGGCCCTGCTCAATATTGACCTTCA 444
Qy 351 LysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGly 370
Db 443 AAGCTT-----ACAATGTTCTCAACA 423
Qy 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
Db 422 GAAACAGATTTGAGAAATTAAGCAAAAATGGAATGAGCGAGGGTCTTTGAGAGGATC 363
Qy 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
Db 362 AACATTGACAGTACGGGGTTGATCCGGTATATATCAACGCGCAGAGACAGGTAATA 303
Qy 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
Db 302 GGGCAGATAGCTCTTGTCTACTTCCAGAACCCCATGGGTTTAGAAAAAGTAGGAATACG 243

Qy 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db 242 ATGTTTATAAATACAGTTTAACTCAGGAGAACCCATGATAGGACAGGAGTGAAC 183
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 182 AGAGCAGCAATAAACCCAGGAGCTTTAGAGATGTCAAATGTCAACTTGGCTTAATGATTT 123
Qy 471 ValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
Db 122 GTAGCATGATTTACAACTCAGAGAGGTTTTCAGGCTAATGCTAAGTTTATAACCGTCTCC 63
Qy 491 AspThrMetLeuGlnLysAlaLeuLeuLysArg 502
Db 62 GATGAATGCTGCAGGATTTAGTGAATGAAGAGA 27

RESULT 18

AE004539

AE004539

LOCUS

DEFINITION

Pseudomonas aeruginosa PA01, section 100 of 529 of the complete

genome.

ACCESSION

AE004539

VERSION

AE004539.1

KEYWORDS

Pseudomonas aeruginosa PA01

SOURCE

Pseudomonas aeruginosa PA01

ORGANISM

Pseudomonas aeruginosa PA01

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE

1 (bases 1 to 25022)

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,

Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,

Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,

Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,

Wong, G.K.-S., Wu, Z., Paulsen, I.T.

Complete genome sequence of Pseudomonas aeruginosa PA01, an

opportunistic pathogen

Nature 406 (6799), 959-964 (2000)

JOURNAL

MEDLINE

PUBMED

10984043

REFERENCE

2 (bases 1 to 25022)

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,

Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,

Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,

Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,

Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sauer, M.H.,

Hancock, R.E.W., Lory, S. and Olson, M.V.

Direct Submission

Submitted (16-MAY-2000) Department of Medicine and Genetics,

University of Washington Genome Center, University Of Washington,

Box 352145, Seattle, WA 98195, USA

JOURNAL

3 (bases 1 to 25022)

REFERENCE

Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)

Direct Submission

Submitted (04-FEB-2003) Department of Molecular Biology and

Biochemistry, Simon Fraser University, 8888 University Dr.,

Burnaby, British Columbia V5A 1S6, Canada

JOURNAL

COMMENT

This represents the February 3, 2003 version of the continually updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.

Class 1: Function experimentally demonstrated in *P. aeruginosa*.

Class 2: Function of highly similar gene experimentally

demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
 Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
 Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

```

FEATURES
  source
    1. .25022
    /organism="Pseudomonas aeruginosa PAO1"
    /mol_type="genomic DNA"
    /strain="PAO1"
    /db_xref="taxon:208964"
    complement(69..770)
    /gene="brag"
    /locus_tag="PA1070"
    complement(69..770)
    /gene="brag"
    /locus_tag="PA1070"
    /note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1"
    /codon_start=1
    /transl_table=1
    /product="branched-chain amino acid transport protein Brag"
    /protein_id="AAG04459.1"
    /db_xref="GI:9946986"
    /translation="MLSFDKSVYTKIQALHDVSVVEKGEIVTLIGANGACKSTLL
    MTLGSPQASGSRVEGEVLGPGSTMRKSIADVPEGRVFSRLTVEENLAMGGF
    FTDKDDYQVMDKYLEPFLRKERYQQRAGTSGSGEQMLAIGRALMSPKLLLDGP
    SLGLPIIIIOQIPEIIEQLRRREGVTFLVQANQALKADRAYVLENGRIVMHDTGA
    ALLNPKVRDAYLGG"
    complement(773..1540)
    /gene="brag"
    /locus_tag="PA1071"
    complement(773..1540)
    /gene="brag"
    /locus_tag="PA1071"
    /note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1"
    /codon_start=1
    /transl_table=1
    /product="branched-chain amino acid transport protein Brag"
    /protein_id="AAG04460.1"
    /db_xref="GI:9946987"
    /translation="MSRPILEVSLTMRFGGLLVNGVNLKVEZKOVSVKIPNGAGK
    TVFNCUTGYPQTGGIIRLDGEEIQGLPGHKIARKGVNVTQFONRLFEMTAVENLL
    VAQRHLNTFPLAGLFTPAFRSREAMEYAAHLEEVNLTFAFRSAGTLAYGQQR
    RLEIARQWTRPRILMLDEPAAGLNPKETDLDKALKAKLSEHNVTLLIHDMDKLVM
    SISDHVINGQAPLADGTPQIQRNDPDVIAKLGEA"
    complement(1537..2790)
    /gene="brag"
    /locus_tag="PA1072"
    complement(1537..2790)
    /gene="brag"
    /locus_tag="PA1072"
    /note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1"
    /codon_start=1
    /transl_table=1
    /product="branched-chain amino acid transport protein Brag"
    /protein_id="AAG04461.1"
    /db_xref="GI:9946988"
    /translation="MSQSLKRALFSAALLVIIVSYPIGLKLRVGIKLEVLGADATQL
  
```

```

WTIAAALAMFVWQLPRDRIPLKIGRGVKGKNGKLNFLSLPSTQRMVAVLALVVVA
FWPFPSRAGVADIATILIVYMLGINTVWGLGLDLGYGVGFVAVGATYVALLAE
YAGGFWTALPACGMALFGLLGPVPLRGLDYLAVTLGFGELIRILLRNMTET
GPNNGISIKPPTLFGLTERRPEGMOTHEFGIAYNTNYKVILLVVALLLVLA
GLVNRMRMPIGRAWELRAEDVACALGINTPIVKLSAFTIGASAGAGSFFAAR
QGLVTPSPFTPIESAMILAIIVLGGMGSLGVILAAVWVLLQMRGNEFRMLIFGL
TMIWMIRPQGLLPMQRPHELEKP"
complement(2787..3710)
/gene="brad"
/locus_tag="PA1073"
complement(2787..3710)
/gene="brad"
/locus_tag="PA1073"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1"
/codon_start=1
/transl_table=1
/product="branched-chain amino acid transport protein Brag"
/protein_id="AAG04462.1"
/db_xref="GI:9946989"
/translation="MPEIYHVLOLVNGLTVSTYALIAIGTVVYGIIGMINFAHGE
VYMGSYIARTAITLLAMGLDSVPLMLAAFAASIIVTSAGVSIERYAVRPLRGN
RLIPLISAIGMSIFLQNAVMSLQSKKAIPTLLPGNFVGESSMNGVIVSYMQLIF
VTVLVMFGLTFLISRSLGRACACAEEDUKNTLLGINSINLIATLTFVIGALAAVA
AVLLMGQVGINPGIFLAGIKAFKTAALVGGISIPGAMLGILLGLLVAREAFGADVFD
QYKQWAFGLLILVLLFRPTGILGRPEVEK"
complement(3971..5092)
/gene="brac"
/locus_tag="PA1074"
complement(3971..5092)
/gene="brac"
/locus_tag="PA1074"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Periplasmic; Subcellular localization confidence: Class 1"
/codon_start=1
/transl_table=1
/product="branched-chain amino acid transport protein Brac"
/protein_id="AAG04463.1"
/db_xref="GI:9946990"
/translation="MKKGTORLSELPAAALTAGFASYSMAADTKIALAGPVTGPVVAQ
IGDWKAGALMAIEQINKAGVGAQLEGVYDDACDPKQAVANVAVNDGVKVFVG
HVCSSSTQPADTIYEDSGVLMTPSATPEITTSRGYKLIPTRTIGLNNQGPVAGKFA
ERYKDKTIAVLHDKQQYEGEGATEVTKTVEDAGIKAVFGLNAGDKDFNALLSKLK
AGQVQVYFGYHPGEMGLLQAKOGLDARFMGPEGVNSEITAIAGDASEGMLATLP
RAFQDPQKALIDAFKAKNQDPSGIFVLPAYSATVIKIEKAGEADPEKVAEALR
ANTETPTGNLGEKDEKDKNFDFTVYVHKKDATRTEVK"
5391..5705
/locus_tag="PA1075"
5391..5705
/locus_tag="PA1075"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG04464.1"
/db_xref="GI:9946991"
/translation="WTEQTSLYAKLLGETAVISWAEIQPFPPFARGALLQVDAALDLVE
VAEALAGDDREKVAAMLGGGLSKVGDAKDFLERDPTLWAVVAVPWVVIQERAEKA
TLH"
5775..6137
/locus_tag="PA1076"
5775..6137
/locus_tag="PA1076"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no
  
```

Alignment Scores: 7,75e-24 Length: 25022
 Pred. No.: 527.50 Matches: 151
 Score: 45.84% Conservative: 86
 Percent Similarity: 29.21% Mismatches: 205
 Best Local Similarity: 20.40% Indels: 75
 Query Match: 1
 DB: 16

US-10-009-823a-1 (1-502) x AB004539 (1-25022)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 DB 8000 AGTTTCAACATCGGCTGAGCGGCATCCAGCGGCTCTAGCGGCTGAACGTCACCGCG 8059

QY 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
 DB 8060 AACAAATCGCCACCGCGCCGAGGCTTCAGCAATCCCGCGGAGTTCGCCGAC 8119

QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 DB 8120 GTCTACGCGCTCGGTG--CTGGGTTCG-----GGCAGCAACCGCGAGGCG 8164

QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 DB 8165 AGCGGGGTGCTGCTCGGACGCTCGCAGATGTTCAACAGGCGCAACATCGACTCGACC 8224

QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
 DB 8225 AACAGCGTGTGACCTGGCCATCAACGCGCAACGGCTTCTGTCACCAACCAACGCG 8284

QY 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
 DB 8285 GCGATCAGCTACACCGCGCGGCTACTTCAATACCGAAGCAGGATTTCATGTCGAC 8344

QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
 DB 8345 AACACGGCTACCGCTCGCGGCTATCGCTCGGCGCGNACCGCGCTGCGAAGCGC 8404

QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 DB 8405 GTGGTCACCGACCTCAAGTCGAGCGCGCAATCAGCGCGCGCGCACTCGACATC 8464

QY 163 LysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGln 182
 DB 8465 CAGAGTCGTACACCTCACTCACTCGCGTGAAG-----CCG 8500

QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
 DB 8501 CCGACCGTGCAGCGGCTC-----GATCGG 8524

QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsn 222
 DB 8525 TCCGACCGCTACCTCACTCACTCGCTTCTCGCTGGGCTATGACAGCGCGGCAAC 8584

QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
 DB 8585 TCCACACCATGACCGAGTCTTTCATCAAGACGCGCGCGCACTGCGACCGCGCG 8644

QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
 DB 8645 ATTCCGAGAGACGCTGACCATGAAAGTGTGATGACGCGGCTCAATCGCGCTCGATCG 8704

QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 8705 TCGAAGACGCGCG-----ATGAGCTTCAACGTCACCTTC 8740

QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 DB 8741 GACGCCAGCGCCAGATGACCTCGGTTCGGCG-----CCGACGCGCAGCAGCGCGG 8794

QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304

DB 8795 CCGGGCTTCAGCATCGACGCGCACCAACAGTGATTCAGTTTCAGTCGGCC----- 8845

QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324

DB 8846 ACTGGCAATCGCGGACTCC-----GGCACCGCTGGATTCGGCGGCGCTCG 8893

QY 325 AspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAa 344

DB 8894 GAC---GGCAAGACCCCGCGACCTACGCTGGAACGCGCGACCGGTGCCGACGGCG 8950

QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364

DB 8951 ATCTCTTCGACATCGCGCAAG-----ACCACCCAGTAC 8983

QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384

DB 8984 TCCACCGCTTCGCCCGCAGACCAACCGATC-----CAGGACGCTACACCAAC 9031

QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404

DB 9032 GGTGATGGCAGCGCTGGAATCGACACACCGGGTGATCTTCGCCCGCTACACCAAC 9091

QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424

DB 9092 GCCCAGTCCAGGTCAGGCGCAGTGTGTGCTGCCCACTTCGCCCAACATCCAGGCGCTG 9151

QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444

DB 9152 AGCGCGATCGGCAAGACCTCTCGGTGAGTCTCTCGGAGTCCGCGAGCGCGCGCTCGGC 9211

QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464

DB 9212 CGCGCGCTCGGCGCACCTCGGGCGCTTGCATCCGCGCGCTGGAAGCGTCCCAACGTG 9271

QY 465 AspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSer 484

DB 9272 GACATCTCCAAAGCACTGTGTGAACCTCATCTCCACGCGCAACTACCGAGCCCAATGCC 9331

QY 485 LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501

DB 9332 AAGACCATCCAGACGAGGATCGCGTGACCCAGACCATCATCACTCGCG 9382

RESULT 19
 BB043739 34817 bp DNA linear BCT 24-JAN-1996
 LOCUS Borrelia burgdorferi fasmid clone 31, complete sequence.
 DEFINITION
 ACCESSION U43739
 VERSION U43739.1 GI:1165250
 KEYWORDS
 SOURCE
 ORGANISM
 Borrelia burgdorferi (Lyme disease spirochete)
 Borrelia burgdorferi
 Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 Borrelia burgdorferi group.
 REFERENCE 1 (bases 1 to 34817)
 AUTHORS Dunn, J.J., Butler-Loffredo, L., Kieleczawa, J., Medalle, J. and
 Luft, B.J.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1995) John J. Dunn, Biology Department,
 Brookhaven National Laboratory, PO Box 5000, Upton, NY 11973-5000,
 USA
 FEATURES
 Location/Qualifiers
 1..34817
 /organism="Borrelia burgdorferi"
 /mol_type="genomic DNA"
 /strain="B31"
 /db_xref="taxon:139"
 /map="approx 290-325 kbp"
 /clone="31"
 complement(1..268)
 /note="orf38; N-terminal portion; Method: conceptual
 translation supplied by author."
 /codon_start=1
 /transl_table=11

CDS


```

MIMLPVWISLFPKILFVMDGWTLLTSLIKSPM"
complement(9847. .10473)
/gene="fliZ"
complement(9847. .10473)
/gene="fliZ"
/note="flagellar protein required for flagella formation"
/codon_start=1
/transl_table=11
/product="FliZ"
/protein_id="AAA85600.1"
/db_xref="GI:1165261"
/translation="NNKFLVFLVFFNFYVANSOVSASDLENESSLPFER
DKANFVNDSTQPSVFINISDLKIVFLIAFFFLKLIIFYSKSKYEQNSLI
KAKYFIEDVKNRIINLDNVIFLVSSNSFTLLKREIKSEEELEDKURLSKINDS
AKKDSFQIFKKYLLKKEIPSSGNDYVKLEKIEASLKQDKRLKP"
complement(10466. .10807)
/gene="fliN"
complement(10466. .10807)
/gene="fliN"
/note="flagellar switch protein"
/codon_start=1
/transl_table=11
/product="FliN"
/protein_id="AAA85601.1"
/db_xref="GI:1165262"

Alignment Scores:
Pred. No.: 1.82e-22 Length: 34817
Score: 509.00 Matches: 150
Percent Similarity: 41.62% Conservative: 71
Best Local Similarity: 28.25% Mismatches: 190
Query Match: 19.68% Indels: 120
DB: 1 Gaps: 13

US-10-009-823A-1 (1-502) x BBU43739 (1-34817)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 15670 ATGATGAGCTCTTATATTCTGGTCTTCTGGCTTCAGAAATCATCAACAAAGATGGAT 15611
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 15610 GTTGTGGTAAACAATATCCCAATGTAATCAATGGCTTTTAAAGAGGAGNATAAT 15551
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer 55
Db 15550 TTTCAGATATGATATCGCAATCTATTCTGGAGCTTCTCGCCCTACTGATCGTGGT 15491
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 15490 GGGACTAATCCCAAGCAAGTTGGATTAGGCATGAATGTTGCCTCAATGACATATTCAAC 15431
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 15430 ACTCAAGGAGCTTTTCAAGCACTCAAAAGCATCTGATCTTGGAGTTAGTGGCAACGGA 15371
QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 15370 TTTTATTATTTAAAGAGGTAATAAATTTGTTTATACAAGAGCGGCTCTTTGACGTG 15311
QY 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
Db 15310 GACTCTGATGACATCTTGTAAATCTCGCAAT 15263
QY 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 15262 GGTGGATGGCAAGAGATTAGAGGTTAAAGGTTATAAATACAGCTTCTGATATTGAG 15203
QY 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 15202 GATCTGATTATTCGATTCGAGATAAAGAGGAGCAAGCTACACGAAAAATGTTACTTTT 15143
QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190

```

RESULT 20
AB001137/c

LOCUS AB001137 44380 bp DNA linear BCT 17-MAR-2003
 DEFINITION Borrelia burgdorferi B31 (section 23 of 70) of the complete genome.
 ACCESSION AB001137 AB0000783
 VERSION AB001137.1 GI:2688160
 KEYWORDS
 SOURCE
 ORGANISM
 Borrelia burgdorferi B31
 Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 Borrelia burgdorferi group.
 REFERENCE
 1 (bases 1 to 44380)
 Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.,
 Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,
 Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D.,
 Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S.,
 Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.,
 Weidman, J., Uterback, T., Watthey, L., McDonald, L., Artiach, P.,
 Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K.,
 Hatch, B., Smith, H.O. and Venter, J.C.
 Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*
 Nature 390 (6660), 580-586 (1997)
 JOURNAL
 MEDLINE
 PUBMED
 98065943
 9403685
 REFERENCE
 2 (bases 1 to 44380)
 Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A.,
 Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,
 Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D.,
 Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J.,
 Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D.,
 Gocayne, J.D., Weidman, J., Uterback, T., Watthey, L., McDonald, L.,
 Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K.,
 Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
 Direct Submission
 Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 Source
 1. 44380
 /organism="Borrelia burgdorferi B31"
 /mol_type="genomic DNA"
 /db_xref="taxon:224326"
 134. 1516
 /gene="BB0261"
 134. 1516
 /gene="BB0261"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="B. burgdorferi predicted coding region BB0261"
 /protein_id="AAC66691.1"
 /db_xref="GI:2688209"
 /translations="MLSLFTIVSSLVFLVFLVFLFKIALKLGIKNGIKKIDRTR
 KLIERALSILKTPNIGALEILNYYIKKDHENGKIVAKLQCLIEDNPINQSEINT
 FRALSYGNNKVRNREALEYRKAFILKKTDEDANYGLVAFKEMYEALYSLRM
 KVVHFNKNNKVLKHLGITFLNLSYQRAVIFNNIKNYIGDIEALIAVAKSLKM
 NOHLALEIKQKQDMGYEALLTSIHSKNKLEKLEONKEIKVDPDPKPT
 FLFLVNLGLOISVYOKQYEAFTKVEIDPNYKNIKEKLEFSLRNLENALRIYL
 RSKENPERKANLILKLINKFQIRDSKINEVTFQIDINFLANNQWENLIVRFV
 RTQDTFGELFLDFISKENKIGLCIAPSKFSLKAKQMEGRLLIDVEGKLAQI
 LKKINISKYT"
 complement (1523. .2776)
 /gene="BB0262"
 complement (1523. .2776)
 /gene="BB0262"
 /note="similar to PID:1652679 percent identity: 36.49;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC66685.1"
 /db_xref="GI:2688203"
 /translations="MIIPKKORVERKKKFLFNSKSVNPELKDFANISIGKRRKK
 VFIKKFFKKKINFFKKVSLFPFKLQINHYEYKYYSLSLRDKDFDIFSVPDYKL
 DCHTDSVGNVCVNVKVGNDNYLLSFPVKKNKEKVMLSLYLPDKFDPKIGKNVNIK
 VFKLNAIFILFTFYINIFSYGVSFVFLNRLSLPKDYFIDTFLYISDQDAIQISSYL
 PESNVASPGFKKNFVLKVDHKKIPEGETLSHVAARYQITSETLSFNEIKDVRNIK
 PNVSVKPNKGIIVYIVKNDSSISSIASAVNPKVDILDDNNLDNVLFLGQKLFIPG
 GRPKDFLKEVLGETFYIPVQGVITSGYGPDPFTGIVSFHNGIDIANLANTPIKAS
 REGVVVTAGNAGGYGKIYIVISHNSNGFQTLIAHLNSFAVKGVKKVSGAIVGNGSTG
 YSTGNHLHFTIFKNGKTENPMKYLR"
 2838. 3344
 /gene="BB0263"
 2838. 3344
 /gene="BB0263"
 /note="similar to SP:P32697 GB:X52093 GB:X54933 PID:47757
 percent identity: 29.61; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="signal peptidase I (lepB-3)"
 /protein_id="AAC66684.1"
 /db_xref="GI:2688202"
 /translation="MRIFKAKHYELASILAACLFLTLIKLFLSYIVKSGEMTPAIF
 EKWIVNHKPAYGLRIKKHQKYLKLLKTKPQNMELIKDPTIKAIKKIIPAIPGEKF
 KQIEKNTICIDHLNFKIDENILKNTKKIPEDHYLVIGENKQISLSDRYGFIKIDNI
 LGKIYYF"
 complement (3370. .4845)
 /gene="BB0264"
 complement (3370. .4845)
 /gene="BB0264"
 /note="similar to GB:M96847 SP:P28608 GB:M97912 GB:S42385
 GB:X67646 percent identity: 40.88; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="heat shock protein 70 (dnaK-1)"
 /protein_id="AAC66683.1"
 /db_xref="GI:2688201"
 /translation="MFMKWIGIDLGTNTVASFVDSKIIILNERGITPSIVSFS
 DKDLVGSAAKQNLINPQFTFYNFKNISNNFYKVDGEFYKAEYISAHLLASVKNR
 AEKFLDEIENAVITVPAYFSEIQRVCVEAANFAGLKAILNEPTAAIAVAFERQ
 IDGIFLYDLGGGTFDVTLMKQGDVTYVLSVQSGRLGNDNFNIEKVLDSPKNG
 YPDFNLEDIFLLQDRRIEIEGKKNLSVMEVDITLLFLDKHLNVLKDEFSNLS
 EYDITQLMEFASLDSDSVISVKIILSGGSTRIPLIEKVLKESFPFSAIIDLALNQ
 DEVAIGAGIHAFSLSRNDSVIRKFDVTPSLGLEIKONEFFTLIKNTALPISRSKL
 FTTNDYQDEIETIHLQEKYKASLANSIGRFSFGNIQKALGKPIEVFLTLNESGI
 LSVNAKLETNSSNFIEIKITSSSDNVAKESLSNTFTSIID"
 complement (4817. .5338)
 /gene="BB0265"
 complement (4817. .5338)
 /gene="BB0265"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="B. burgdorferi predicted coding region BB0265"
 /protein_id="AAC66690.1"
 /db_xref="GI:2688208"
 /translation="MFIVFYILIIIFIFVHYVINLKKSNILKRFKSEVDKTIIE
 INQATDNINIIEIKESLNRIKEVDRIEIKDLQRLGNSSIPGNSPSSGSKSD
 GIKYNSLDYSMPTIEKNIKESYNVRNRIILHEQGMFSAIAKFKLDLGEVELIIS
 IHRGVSVEKMDRY"
 complement (5358. .5666)
 /gene="BB0266"
 complement (5358. .5666)
 /gene="BB0266"
 /note="similar to PID:1165251 GB:AE000783 percent
 identity: 100.00; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC66682.1"
 /db_xref="GI:2688200"

gene complement (5687. .7591)
/gene="BB0267"
CDS complement (5687. .7591)
/gene="BB0267"
/note="similar to PID:1165252 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC66681.1"
/db_xref="GI:2688199"
/translation="MDDRINFSELCDKMSRYLERSHVDLMEIADTLEBALNDASL
ELSPVYKELDYKILGNGILGKQKWKIVAYRNSKSPKVSIDNSDSEKEVIS
SDGEPIFKRSKGVFLKVIKSPKGQIPKIQDVIKLELHNSNIENFNKDIKVGIVENA
SCYSEKIDAFESDPENVMVQISENSMSVTIETPAINGAEIFAODIENILKRYG
VIDIALKDKVAKFYDVPYKGPVELARGLKPKDSYIDFLVDENNRFGYEVSSI
GRNTEGELARIIPFGKVDGYTVFGKLVKSKESQDIDFILCENTFRDGYKIRAKC
LYGMSVDGIVSHNIIYKVGVDGPATGDIVNNGMVLQGVNLDGINIMAKSGIEVKG
LVRCNLTGDSGIIIFHSYNGKDSKIYARGNVRKAFLENTLNGCGSEVEVLGIVNS
SYSSKRVLCIKGKSKVCAKEVRAYSIGSDSCETCIDVYDPEIKDLITRF
TSHLEKI ERLVLTAKDI VALKKNIVLITDKAEKSLKIDSYNEFVNESKILRMEIKMI
EAKRLNLONELENTKIEGKSVENIACSGVKLYIKNAFVELSKDYNVNTFIEDDNYIK
VMYIIPFKSR"
complement (7594. .8073)
/gene="BB0268"
CDS complement (7594. .8073)
/gene="BB0268"
/note="similar to PID:1165253 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC66680.1"
/db_xref="GI:2688198"
/translation="WFISRELKVIILTSVSALFISAVLGLFGLSFFTLVRSLLQFV
PFVGLLEIYFKYLSNLFSTELSGNENKPDQDKSDXDFKNLDFONKNSLYEN
SQRISSGGDFIEVKYKFETEDVSRGSDKNKMSFIENDPKVADALKILMSKE"
complement (8080. .8967)
/gene="BB0269"
CDS complement (8080. .8967)
/gene="BB0269"
/note="similar to PID:1165254 percent identity: 99.64; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP-binding protein (ylxH-1)"
/protein_id="AAC66679.1"
/db_xref="GI:2688197"
/translation="VEDQASLRDMMRLNGKFNFSVDEKVNQSKTRFIATVSSGKGVG
KSNIAIGLAKYSELGKVLIIADIGMANVNIILGVIPKYSIYHMIASDRDIRVIT
KTEYNIDLAGASGTMEIIDLSDVDVNRKFIKELLYEYDIIVIDTSAGISKQVLSFL
LSDDVVIIVTPEPTSIDTAYGIIVKLYSYMENLKNLRILNVRNANVSEAGVAKVI

gene complement (5687. .7591)
/gene="BB0267"
CDS complement (5687. .7591)
/gene="BB0267"
/note="similar to PID:1165252 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC66681.1"
/db_xref="GI:2688199"
/translation="MDDRINFSELCDKMSRYLERSHVDLMEIADTLEBALNDASL
ELSPVYKELDYKILGNGILGKQKWKIVAYRNSKSPKVSIDNSDSEKEVIS
SDGEPIFKRSKGVFLKVIKSPKGQIPKIQDVIKLELHNSNIENFNKDIKVGIVENA
SCYSEKIDAFESDPENVMVQISENSMSVTIETPAINGAEIFAODIENILKRYG
VIDIALKDKVAKFYDVPYKGPVELARGLKPKDSYIDFLVDENNRFGYEVSSI
GRNTEGELARIIPFGKVDGYTVFGKLVKSKESQDIDFILCENTFRDGYKIRAKC
LYGMSVDGIVSHNIIYKVGVDGPATGDIVNNGMVLQGVNLDGINIMAKSGIEVKG
LVRCNLTGDSGIIIFHSYNGKDSKIYARGNVRKAFLENTLNGCGSEVEVLGIVNS
SYSSKRVLCIKGKSKVCAKEVRAYSIGSDSCETCIDVYDPEIKDLITRF
TSHLEKI ERLVLTAKDI VALKKNIVLITDKAEKSLKIDSYNEFVNESKILRMEIKMI
EAKRLNLONELENTKIEGKSVENIACSGVKLYIKNAFVELSKDYNVNTFIEDDNYIK
VMYIIPFKSR"
complement (7594. .8073)
/gene="BB0268"
CDS complement (7594. .8073)
/gene="BB0268"
/note="similar to PID:1165253 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC66680.1"
/db_xref="GI:2688198"
/translation="WFISRELKVIILTSVSALFISAVLGLFGLSFFTLVRSLLQFV
PFVGLLEIYFKYLSNLFSTELSGNENKPDQDKSDXDFKNLDFONKNSLYEN
SQRISSGGDFIEVKYKFETEDVSRGSDKNKMSFIENDPKVADALKILMSKE"
complement (8080. .8967)
/gene="BB0269"
CDS complement (8080. .8967)
/gene="BB0269"
/note="similar to PID:1165254 percent identity: 99.64; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP-binding protein (ylxH-1)"
/protein_id="AAC66679.1"
/db_xref="GI:2688197"
/translation="VEDQASLRDMMRLNGKFNFSVDEKVNQSKTRFIATVSSGKGVG
KSNIAIGLAKYSELGKVLIIADIGMANVNIILGVIPKYSIYHMIASDRDIRVIT
KTEYNIDLAGASGTMEIIDLSDVDVNRKFIKELLYEYDIIVIDTSAGISKQVLSFL
LSDDVVIIVTPEPTSIDTAYGIIVKLYSYMENLKNLRILNVRNANVSEAGVAKVI

Alignment Scores:
Pred. No.: 2, 43e-22 Length: 44380
Score: 509.00 Matches: 150
Percent Similarity: 41.62% Conservative: 71
Best Local Similarity: 28.25% Mismatches: 190
Query Match: 19.68% Indels: 120
DB: 1 Gaps: 13

US-10-009-823A-1 (1-502) x AE001137 (1-44380)

Qy 1 MetMetGlySerLeuPheSerGlnAspLeuAlaLeuGlySerHisSerThrGlyLeuGly 20
Db 21060 ATGATGAGGCTTTATATTCTGTTCTGGCTTCAGAAATCATCAAAAGAAATGGAT 21001
Qy 21 ThrValSerAsnLeuAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 21000 GTTGTGTTGTAACATATCGCAATGTAATACATTTGGCTTTTAAAGGAGAGATTAAT 20941

```

Db 20130 TCTAGTAGCACAAGGCTATTATTCAAGATGGATATGCGATGGATATATGGAATAATAT 20071
Qy 391 ThrIleThrSerGluGlyLeuGlnGlyLysTy-SerAsnSerGlnValValAspPhe 410
Db 20070 GAAATTGATCAAAATGGTGTATAGTTGGCATTATTCAATGGCATAAGACGAGATCTT 20011
Qy 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
Db 20010 GCGAAGATGCTCTGCTCTCTTTATGAATCCCGAGGACTTGCATAAATCAGCGGATACT 19951
Qy 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db 19950 AATTTGTAGAACACAGCAATTCAGTCAAGTAGATAGACGAACTGACCTGCTGGA 19891
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 19890 CTTGGTGATATTAGATCTGGTGTGTTTAGAAATGCCCAATGTTGATCTTGCAGAGCAATT 19831
Qy 471 ValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
Db 19830 ACAGATGATAGTAGACCAAGAGGATTTACGCAATGCCAARACCATACCATTCTT 19771
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 19770 GATCAATTATTACAAGAACTGTGAAGATTGAAA 19738

RESULT 21
LOCUS BORFLGE 1499 bp DNA linear BCT 06-FEB-1996
DEFINITION Borrelia burgdorferi flagellar hook protein (flgE) gene, complete cds.
ACCESSION L43849
VERSION 1 GI:1184295
KEYWORDS flagellar hook protein; flgE gene.
SOURCE Borrelia burgdorferi (lyme disease spirochete)
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
Borrelia burgdorferi group.
REFERENCE 1 (bases 1 to 1499)
Ojalini, C., Davidson, B. E., Saint Girons, I. and Old, I. G.
Conservation of gene arrangement and an unusual organization of
rRNA genes in the linear chromosomes of the Lyme disease
spirochaetes Borrelia burgdorferi, B. Garinii and B. afzelii
Microbiology 140 (Pt 11), 2931-2940 (1994)
MEDLINE 95111614
PUBMED 7812434
REFERENCE 2 (bases 1 to 1499)
Old, I. G.
Unpublished (1996)
On Feb 11, 1996 this sequence version replaced gi:903702.
Original source text: Borrelia burgdorferi (individual isolate
HB19) (clone: PAD12B) (clone library: HB19 Cosmid) DNA.
FEATURES
source
1..1499
/organism="Borrelia burgdorferi"
/mol_type="genomic DNA"
/isolate="HB19"
/db_xref="taxon:139"
/clone="PAD12B"
/clone_lib="HB19 Cosmid"
115..1443
/gene="flgE"
115..1443
/gene="flgE"
/codon_start=1
/transl_table=11
/product="flagellar hook protein"
/protein_id="AAA87351.1"
/db_xref="GI:1184296"
/translations="MRRSLYSGVSGLQHQITMDVVGNINAVNTIGPKGVNFQDM
ISQISGARPTDARGTKQVGLGNVASIDTITQAFQSTQASDLGSGNGFF
ILKEGKNLFYTRAGAFYVDSRHLVNPANGMRIQGMWARDLEGEKVINTASDIEDLI

```

```

PIGKEGAKSTNTTTFACNLKRLPLIQEGASPADIAECTTWVNVKSLYDSFGNVSVLE
LRVKDNLNPLNATVLINGEONSFTLGFNEGALASLNGQPGQGDILQIPITEN
VLGANVEVGEQVNLKLTGVSYSYDSITQFADSSSTKAIIDQYGMGMENYEDQ
NGVTVGYSNGIRDLKIALASFMNPGKLAGSDTNFVETSNQVRIGETGLAGLG
DIRSGVLEMANVDLAEOFTDMIVTQRFQANAKTITTSDDLQELVRLKN"

ORIGIN

Alignment Scores:
Pred. No.: 1,08e-23 Length: 1499
Score: 503.00 Matches: 149
Percent Similarity: 41.62% Conservative: 72
Best Local Similarity: 28.06% Mismatches: 190
Query Match: 19.45% Indels: 120
DB: 1 Gaps: 13

US-10-009-823A-1 (1-502) x BORFLGE (1-1499)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 115 ATGATGAGGTCTTTATATCTCTGGTGTTCGGCTTCAGAAATCATCAACAATAATCGAT 174
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValVal 40
Db 175 GTTGTGTGTAACAATATGCCAATGTAAATACAAATTCGCTTTAAAAAGGGAAGAGTAAT 234
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Db 235 TTTCAAGATATGATATCGCAATCTATTTCTGGAGCTTCCTCGCCCTACTGATGCTCGGGT 294
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 295 GGGACTAATCCAAAGCAAGTTGGATTAGGCAATGTTGCCTCAATGTGACATATTTCAC 354
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 355 ACTCAAGAGCTTTTCAAGCACTCAAAAGCATCTGATCTTGGGGTAGTGCAACGGA 414
Qy 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 415 TTTTATTATTTAAAAGAGGTAAAAATTTGTTTATACAGAGCGGTGCTTTTACGTG 474
Qy 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgGlySer 135
Db 475 GACTCTGATGACATCTTGTAAATCCTGCAAA-----GGAATCGCAATTCAA 522
Qy 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 523 GGTGGATGGCAAGAGATTTAGAGAGGTAAAAAGGTTATAAATACAGCTTCTGATATTGAG 582
Qy 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 583 GATCTGATATTATCCGATTCGAGATTAAGAGGAGCAAAAGTCTACTAAAAATATACCTTT 642
Qy 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190
Db 643 GCTTGTAACTCTGATAGAGATTGCCCTTAATTCAGAGAGGTGCGAGTCTCGCAGATATT 702
Qy 191 LeuLeuGluSerThrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
Db 703 GCACGCGGAACTTGGGTGTCAATAAATCA----- 732
Qy 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
Db 733 -----TTGTATGACAGTTTGGAAAT-----GTTAGTGTCTT--- 765
Qy 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
Db 765 ----- 765
Qy 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuMetSerGlyThr 270
Db 765 ----- 765

```

271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
766 -----GAGCTTAGA----- 774
291 AlaThrLysAspLeuAsnAla-----TrpGlnProAlaProLeuValAsnGlyLeu 307
775 GTTGTAAAGATTAAATACGCTAATTATGGAATGCAACAGATTATTAATAATGCT--- 831
308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe--- 326
832 ---GAGCAAAATCAATTT-----ACATTTGGGTTTGAC 864
327 -----GlyIleLysSerGlnGlnAsnMetTrpAlaGly 337
865 AATGAAGAGCAATGGCGCTTTTAAATGGTCAACAGGGCAAAAGAGATATTTCTTCAA 924
338 AlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetPro 357
925 ATTCTTATACATTAAATGTTTGGGTGCAATGTAAGTAGGTGAGTGGTGGTGGTGGTGGT 984
358 IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370
985 GTAAATTTGAAATTTGGAACAGTTGGAGTTACACTGATTCATTAATCACTCAGTTTGTGAT 1044
371 SerSerThrArgArgTySerGlnAspGlyTyProGlnGlyAspLeuValAspVal 390
1045 TCTAGTAGCACAAGGCTATTATTCAGATGGATATGGCATGGATATGGAATAATAT 1104
391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTySerAsnSerGlnValValAspPhe 410
1105 GAAATTTGATCAAAATGGTGTATAGTTGGATTTTCAATATGGCATAAGACGAGATCTT 1164
411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
1165 GGCAGATTTGCTCTCTCTCTTTATGATATCCGAGGAGCTTGCATAATCAGCGCATCT 1224
431 HisTySerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
1225 AATTTGTAGAACAGCAATTCGGTCAAGTAGATAGTGGCAACTGACTTCTGCTGA 1284
451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
1285 CTTGGTGTATATAGTCTGCTGTTTATGAAATGGCCAAATGTTGATCTTCAGACCAATT 1344
471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
1345 ACAGATATGATAGTACCCCAAGAGGATTTACGCAAAATGCCAAACCACTTACCACTCT 1404
491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
1405 GATCAATTTATACAGAACTTGTAGATTGAAA 1437

RESULT 22
AL646078 203050 bp DNA linear BCT 02-SEP-2002
LOCUS Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
DEFINITION segment 3/11.
ACCESSION AL646078 AL646053
VERSION AL646078.1 GI:17430778
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE 1
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Ariat,M., Billault,A., Brottier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choizeau,N., Claudel-Renard,C., Cunac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Signier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Nature 415 (6871), 497-502 (2002)

MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

21681879
11823852
2 (bases 1 to 203050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1..203050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(147..1499)
/gene="RSP0310"
/note="synonym: RS05460"
complement(147..1499)
/gene="RSP0310"
/functions="cell processes; transport of small molecules"
/note="product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CAD17461.1"
/db_xref="GI:17430778"
/db_xref="TrEMBL:Q8XT07"
/db_xref="GOA:Q8XT07"
/translation="MNAPRKNTDRAEAALLRNASTTHEPNPCWIPDAAAAGLHPGPR
WRAPFAFGQALSLGSALTQVLLWMTDTTGSVSALATAGLALLPQAVLSPLG
GTFADRSRLMLIVADGIALCMILLIALFLTGRIALWYVYVMAIRSAQAQAPAA
AASVTMLVPGGFLVRAAGLNSQLSTLVVAAPLGAALGVMPILGALGVLTATALL
GIAPILCCRIPQACVSSGQKTLMPAFEGVDLVWKTPTGLRWLYLVGLVLTIMPTF
TLVPLVTHFGGAAQVAMLMGSLGIGMAGVLLVMAAPRQVAVILCGFAMSCVA
LALTALAPGSLFGVAVAWVTSITTFVGNAPLITLQITVPNHLOQRVLSLNTTNG
LAAPVGLALTPTLGEVIGVRAIPALGALGAAASLAGFLSPSLMRLDRPFRSEARTD
"

complement(1531..2961)
/gene="oprM"
/note="synonyms: RS05459, RSp0311"
complement(1531..2961)
/gene="oprM"
/functions="cell processes; transport of small molecules"
/note="product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABLE OUTER MEMBRANE CHANNEL LIPOPROTEIN"
/protein_id="CAD17462.1"
/db_xref="GI:17430780"
/db_xref="GOA:Q8XT06"
/db_xref="TrEMBL:Q8XT06"
/translation="MELRIIPLYVALSACSLIPALIKPALVPVATFPABSPETQAAH
ADLGWRTWFGDRRLQRLIGLITLANNRDLRLALNVEYVYQYSIQRAARLPGLDQGS

ASRQRTAANDGAPPAALANSEQYGVNVLNAFEIDLFGVRSLSDAAFYARLASDQG
RRAQISLVGADAVFAERLAQEQROLAETHLADWRQSLDLARLRKATKQNSDLDDVA
QAEQVANAQADLEARTALAAQAGNSLQRLVAGLADLPAFLSLETPQVVTQLPAGL
PSELIPRDIQROEHNLAANADIGAARAAFPRLSLTASLGASPAKGSFLDGGHR
VWFSPOITOPILQGGRLRELRLAELRKSSAIAEYERTIQTAFRVADGLAGRATYG
RQIEAQIRAVASRVELNRLRYRAGVGEGRLELDSQRLYASRQALLDLRQGAFSN
AVALYKALGGLEAETDVATDRVARK
/gene="mexD"
/note="synonyms: RS05457, RSp0312"
/gene="mexD"
/note="synonyms: RS05457, RSp0312"
/gene="mexD"
/note="synonyms: RS05457, RSp0312"
/functions="cell processes; transport of small molecules"
/note="Product confidence: probable
Gene name confidence: putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE
PROTEIN"
/protein_id="CADI17463.1"
/db_xref="GI:17430781"
/db_xref="GOA:Q8XT05"
/db_xref="TrEMBL:Q8XT05"
/translation="MPQFFIRRVFAWVIALFILILGAIAIPKLPRIYPSVAPPTVN
LSATPGATQTNWDGVLISIERELSSKNLYFESSSDTSASIVATFREGTDPAM
AQVDQNKLAIEPRIPQVVRQGLTVESAGSGFLMIIGLQSDGDFDETLADYNAR
NVVEELRIRIGVRVOLFGEAQMRITWDPAKLIAGLNDLSAAISOONAQIAPGS
VGSALPAPQRTVPLTQQLTTPQQAIVLVRANSGKVMGVNVARVELGAQSYG
FINREKQAATAAVALQAPANAVTAEGVQARMELSRNPMAGMSISIPFTAFPVK
ISIEKVITLIEAVLVLVFLQNVYTLIPAVIAPALIGFTIMLAGSGFINV
LTFMGVLAIGTIDDAIVVVENVERLMAQELSPDRTAKMKETIGAVIGITVLT
AVFIPMALASGVSVIQQFTSNVSIIVSFAFLTLTPALCATLTKPVDPAHKEK
GVFGVNRDRITQGVAAVGRKLVTRGRVLMVFGAIAAVLALAFQPLSPFLPEED
QGVFMFTQLPADATAEITLDVKAPEEHVASEASIQSNMSIVGFGSGSGSNAMAF
TMKMDPARQATQNEMLAQQAQAGATEGTVNLSLIPALDELGTSSGFTLRQDRA
NQDAAALAEKTLLEAAQSRISVGYDPLGATGSIKIDRQKAEALGVAFTSIS
ETLSATGSLVYDFNAGRMQVVIQADAPARMQIDDLKLYVRNAGSGVPLSEW
RPVMSPTQLQVRFQGPAPARIIGNAAPGASGAAMERLAAQLPPGFATAWTQS
LQERQSAQAQPMALMSLVVLVLAALYESWALPLSVMLVPLGLGVALVLRGM
PNDVFFKVGMITVIGLSAKNAILIVFAKOLRQEGKGLVESAVTAARLRILPILMTSL
AFGLGVVPLMIANGASATQHAIGTVGFGVGSWATVLAIFFPVFPVFWMLQERPEA
WRASRQAPPPFPVHQFQEG
/complement(6136..7293)
/gene="mexC"
/note="synonyms: RS05456, RSp0313"
/complement(6136..7293)
/gene="mexC"
/function="cell processes; protection responses;
drug/analog sensitivity"
/note="Product confidence: probable
Gene name confidence: putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM PROTEIN"
/protein_id="CADI17464.1"
/db_xref="GI:17430782"
/db_xref="GOA:Q8XT04"
/db_xref="TrEMBL:Q8XT04"
/translation="MKIERLRAGGVHVLAAVLACGSEPSQSPAPPAEVAATTVPAP
AHLEETDLRGRVAARVAREIQVSGIVORRLFEQTEVRAGQELFQINPAFPRA
DTAAALRLAEALRARVOTTLQPLVEDADVSRQADVDSQDQAAADVEQARAT
LAARQLKPAATEAPIPGRIDQALVTEGALVSSGDSNPARIQIDQVYVWRFPAS
LNRQLATQKTGADGFLPVAIRNNGEPYKAGHILFSGINVDAGTGVDLLRVLV
NNPQRLPLQMWVRARVPHARYTDALTVPQAVRVVGGQPOVMTLIDANSHARKLPVEL
GELADRRYRIRIAGLSAQKIVSGMERLSDGAAVAADHDKSPPEVPAGPAPR"

7392..8162
/gene="RSp0314"
/note="synonym: RS05454"
7392..8162
/gene="RSp0314"
/function="miscellaneous; not classified regulator"
/note="Product confidence: probable
Gene name confidence: hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE TWO COMPONENT RESPONSE REGULATOR
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CADI17465.1"
/db_xref="GI:17430783"
/db_xref="GOA:Q8XT03"
/db_xref="TrEMBL:Q8XT03"
/translation="WEINORGAAAVVPMPPSGAQAQALIIAEDPEIADILAA
YLAASGURSVAADGREALEHLKSLKPDLLVLDVQMPHVGWKALESIRHGRDTPVM
LTALDQDIDKLMGLRIQADYVVKPFPNPAVARTQVLRSSMSGTERRDQVLRVAP
FQIDLNHEATVQVQKORHALALTLEFKLLAQLARAPKRVFRAELLAACLPEGDTL
ERTVDSHISIKLRKKLEDLGVNGMPVGVGVYRLWSLE"
8159..9241
/gene="RSp0315"
/note="synonym: RS05453"
8159..9241
/gene="RSp0315"
/EC_number="2.7.-."
/function="miscellaneous; not classified regulator"
/note="Product confidence: probable
Gene name confidence: hypothetical
predicted by Codon usage
predicted by Framed"

Alignment Scores:

Pred. No.:	4,2e-21	Length:	203050
Score:	502.00	Matches:	137
Percent Similarity:	44.40%	Conservative:	85
Best Local Similarity:	27.40%	Mismatches:	168
Query Match:	19.41%	Indels:	110
DB:	1	Gaps:	15

US-10-009-823A-1 (1-502) x AL646078 (1-203050)

QY	8	GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla	27
DB	44995	GGCCTGACGGCTCTCGACGGCGCTCGAAGAACCTGACGTGCGGCAACGTCGCC	45054
QY	28	AsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln	47
DB	45055	AATGCCACACGGTCGGCTACAGAGTCGACGGCCGAGTTCGGCAGCGTATGCGCGT	45114
QY	48	AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln	67
DB	45115	TCGCTG---GTCGGCGCGCTCC-----GACAAACAGATCGCGCGCGTGGAC	45159
QY	68	ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr	87
DB	45160	GTGACCAAGGTGTCGAGTCGTCACCGAGGCAACGTGACCGTACCGGCAACCCGCTG	45219
QY	88	AspLeuAlaIleGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysVal	105
DB	45220	GACATCGCATCAACGGCACCCTTCTACCGATGTCGACGCGTCCGAGCGCGAGGTG	45279
QY	106	HisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer	125
DB	45280	TCGTACACCCCAACGGCAGTTCACACCGACGAAGAACGGGTACATCATCTCGCCACC	45339
QY	126	GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeu	145
DB	45340	GGCCAGAACCTGACCGGCTAGCGGCTGCGACCGCGCAAGATCAACACCGCGCTGCTG	45399

QY 146 GluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerPro---AlaLysThr 164
Db 45400 ACCAACTCGCATGATCCGGTGAACGAT-----CTGCGCGCGTGGACACAC 45447
QY 165 SerThrAlaLeuAsnAlaValAlaLeuGlyAspSerThrAspLysThrGlnSerGlu 184
Db 45448 AACACCGCGTTCTCC-----ATCAACCTC---GATCGCGCGGACGCGTGGCGAC 45495
QY 185 AlaAsnProTyPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProTyrSer 204
Db 45496 -----TyrSerTyAlaGlnProMetArgValTyrAspGlnGlnGly 45510
QY 205 ThrSerAsn-----TyrSerTyAlaGlnProMetArgValTyrAspGlnGlnGly 221
Db 45511 GCCACCACTCGCCACGTCACCACTCGGTGTCGACAGCGTGTACAGCGGACGGC 45570
QY 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241
Db 45571 ACCTGCGACATGTCACCACTATTAC-----GTGCGCACGCGCGCGTGGATGTG 45624
QY 242 PheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAsp 261
Db 45625 TACTCGCAGGTGACCGCGCAATCGACCGCGGCAACCGGTACGTCG----- 45675
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281
Db 45676 -----CTGACCTTCAACGACGACGCGGTGCTG----- 45702
QY 282 MetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAla 301
Db 45703 -----ACCTGTCGCGGACGACGAGTGGCGTGGCGCTTC----- 45735
QY 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
Db 45736 -----ACCGGACAC-----CGCGGACGATGACATC 45750
QY 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProLaser 341
Db 45751 GCCAACATGACACTC----- 45765
QY 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
Db 45766 -----ACCGGACAC----- 45774
QY 362 SerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGly 381
Db 45775 -----ACGAGTACGCGCGCGGCTTCAACGACACACACCGCTCTCGCAGGACGCG 45822
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 45823 TACGCCACCGCGCGCTGCGCACCTACTCGTGGGACCGGATGGACCATCAACGCGCGCG 45882
QY 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 45883 TACTCCACCGCGCGCACGTCGCGCGGATCGCCATGCCACCACTCAAGGCGCGCG 45942
QY 422 AspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyPro 441
Db 45943 GACGCGGTGAGATATCGCGCGCAACGATGGGTGGAGACCGCGGATCGCGCGCGCG 46002
QY 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 46003 CAGATGGGACCGCGGCGCATGGCTGCTCGGCTGCTGCGTGGCGGTGGAGCAG 46062
QY 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGln 481
Db 46063 TCCACGTGGACCTGACGCGCGGAGTGGTCAACATGATCGTCGCGCAGCGCTGATACAC 46122
QY 482 MetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 46123 GCCAACGCGCAGACCATCAAGACCGGAGACCATCTCTCCAGACCTGCTGTGACGCGC 46182

RESULT 23
STFLGE
LOCUS 1320 bp DNA linear BCT 08-MAR-2000
DEFINITION Salmonella typhimurium flgE gene for flagellar hook protein.
ACCESSION X51737
VERSION X51737.1 GI:47671
KEYWORDS flagellar hook protein; flagellar protein; flgD gene; flgE gene;
flgF gene.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 1320)
AUTHORS Homma, M., DeKosier, D.J. and Macnab, R.M.
TITLE Flagellar hook and hook-associated proteins of Salmonella
typhimurium and their relationship to other axial components of the
flagellum
JOURNAL J. Mol. Biol. 213 (4), 819-832 (1990)
MEDLINE 90294298
PUBMED 2193164
REFERENCE 2 (bases 1 to 1320)
AUTHORS MacNab, R.M.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1990) Macnab R.M., Yale University, Dept of
Molecular Biophysics and Biochemistry, Box 6666, 260 Whitney
Avenue, New Haven CT 06511, U S A
COMMENT See <X52094> for flgF gene.
FEATURES
Location/Qualifiers
1..1320
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
/map="23min"
<1..34
/codon_start=2
/transl_table=1
/product="flgD gene product (11 AA)"
/protein_id="CAA3602.1"
/db_xref="GI:809769"
/db_xref="GOA:P16321"
/db_xref="Swiss-Prot:P16321"
/translation="TTLDEVQRQII"
61..1272
/codon_start=1
/transl_table=1
/product="flgE protein (AA 1-403)"
/protein_id="CAA3602.1"
/db_xref="GI:47673"
/db_xref="GOA:P16322"
/translation="MSESOAVSLNRAATLVIGNNIANSATYGFKSGTASPADMEA
GSKVLGVKAGITQDFIDGITITNGRLVAISQGFRLVDSNGSYFYSRKGQFKL
DENRLVNMQGMQLTGYPATGPTIOOGANPAPITIPNTLMAKSTITIASMQNLNS
TDPVSKTPFSVDADSYNKKGTVTYVDSQNAHDMNVYFKTKDNEMAVYTHDSDP
AATAPTASTTLKFNENGILSGGTGVTNITGTGATAATFSLFLNSMQGTGANNI
VATNONGYKPDILVSYQINNDGTGVGNSYNEQVQLQIVLANFANNEGLASQGDNVW
AATCAAGVALLGTAGSGNFGKLTNGALEASNVLDLSKELVNNIQAQRYQNAQTIKTQ
DQILTLVNLK"
1293..>1320
/codon_start=1
/transl_table=1
/product="flgF gene product (AA 1-10)"
/protein_id="CAA3603.1"
/db_xref="GI:581759"
/db_xref="GOA:P16323"
/db_xref="Swiss-Prot:P16323"
/translation="MDHALITYAM"

Alignment Scores:
Pred. No.: 1,96e-23 Length: 1320
Score: 498.00 Matches: 144
Percent Similarity: 43.91% Conservative: 76

Best Local Similarity: 28.74%		Mismatches: 167	
Query Match: 19.26%		Indels: 114	
DB: 1		Gaps: 14	
US-10-009-823A-1 (1-502) x STPLGE (1-1320)			
QY	10	ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla	29
DB	82	AGCGGCTGAACGCTGGCGCCACCAACCTGATGTATCGGTAATAACATCGCCAACTCC	141
QY	30	AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu	49
DB	142	GCACCTATGGCTTAAGTCCGGTACGGCATCTTGGCGATATGTTCCGC-----	192
QY	50	AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly	69
DB	193	-----GGTTCC-----AAAGTGGGCTGGCGGTAAAGTGGCG	225
QY	70	SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu	89
DB	226	GGGATTACCCAGGATTTACGACGGTAGCAGACCAACACCGCGCGGGCTGGATGTC	285
QY	90	AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr	108
DB	286	GCGATTAGCCAGAACCGTTTTCGGCTGTGTAGACAGCAACCGTTCGGTGTCTATAGC	345
QY	109	ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr	128
DB	346	CGCAACGGCCAGTTCAACATGCGACGAAACCGTAACCTGTGTCATATATGCGAGGATGCAG	405
QY	129	LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle	148
DB	406	TTGACGGGTATCCGGCCACCGGTACGCGCGCGACCATTCAGCAGGGGGGAATCCTGCG	465
QY	149	GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
DB	466	CGGATCACCATTTCCGAACACGCTGATGGCG-----GCGAAATCGACCAACCGCG	516
QY	169	AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
DB	517	TCATGCGATCACCTG---AACTCAACGACCCCTGTACCTCTAAACG-----	564
QY	189	PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPheSerThrSer-----	206
DB	565	-----CCCTTTAGCGTGAATGATGCG	585
QY	207	---AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp	225
DB	586	GATTTCGTATACAAAAGGACCGCTACCGTTTATGACACCGGGTAAATGCCCATGAC	645
QY	226	IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu	245
DB	646	ATGAACGCTCTATTT-----GTGAAACCAACCAAGATAATGAATGGGCC	687
QY	246	ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu	265
DB	688	GTGTACACCATGACGACGCGATCTCTGACGCCACTGCGCCAAACACGCGC-----	738
QY	266	LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys-----	280
DB	739	-----TCCACTACGCTGAATTCATGAAACCGGATTCGTGAGTCTGGCGGTACGGTG	792
QY	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro	300
DB	793	AACATCACCAACCGGTACCATTAATGGCGCGACGCG-----	828
QY	301	AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln	320
DB	829	-----GCCACCTTCTCCCTCAGCTTCCTTAACCTCC-----	858
QY	321	ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla	340
DB	859	-----ATGCAGCAGAAC-----	870
QY	341	SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr	360
DB	871	-----ACC	873
QY	361	SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp	380
DB	874	GGGCTAATAACATCGTCGCCACCAAT-----CAAAAC	906
QY	381	GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly	400
DB	907	GGCTATAAGCCGGCGACCTGGTACGATTAACCAACGACGCGCACCGTGGTGGC	966
QY	401	LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer	420
DB	967	AAC TACTCAACGACGACGACGAGTCTGGGCGAGATTGTGTGGCTAACTTCGCCAAC	1026
QY	421	GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly	440
DB	1027	AACGAAGTCTGGCATCCAGGGCGATAACGTCCTGGCGCGACGACGAGGCTTCGGGGTT	1086
QY	441	ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu	460
DB	1087	CGCTGTCTGGGACTCGCGTTCGGCAACTTCGTAAGCTGACGAAACGCGCGCTCGAA	1146
QY	461	ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe	480
DB	1147	GCCTCTAAGTGGATTTCAGTAAGAGCTGGTGAATATGATCGTCGCGCAGCGTAACTAC	1206
QY	481	GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu	500
DB	1207	CAGTCGAATGGCAGACCATCAAAACCCAGGACCATCTCAATACGCTGGTTAACCTG	1266
QY	501	Lys	501
DB	1267	CGC	1269
RESULT 24			
AE008751			
LOCUS			
DEFINITION			
Salmonella typhimurium LT2, section 55 of 220 of the complete genome.			
ACCESSION			
AE008751.1 Gi:16419687			
VERSION			
AE008751.1			
KEYWORDS			
SOURCE			
ORGANISM			
Salmonella typhimurium LT2			
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.			
REFERENCE			
1 (bases 1 to 20753)			
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,			
Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,			
Hou, S., Layman, B., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,			
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, I., Miller, W.,			
Stonking, T., Nhan, M., Waterston, R. and Wilson, R.K.			
Complete genome sequence of Salmonella enterica serovar Typhimurium			
TITLE			
LT2			
JOURNAL			
Nature 413 (6958), 852-856 (2001)			
MEDLINE			
21534948			
PUBMED			
11677609			
REFERENCE			
2 (bases 1 to 20753)			
AUTHORS			
The Salmonella typhimurium Genome Sequencing Project			
CONSTRM			
Direct Submission			
TITLE			
JOURNAL			
Submitted (29-MAR-2001) Genome Sequencing Center, Department of			
Genetics, Washington University School of Medicine, 4444 Forest			
Park Boulevard, St. Louis, MO 63108, USA			
COMMENT			
Supported by NIH grant SU 01 AI43283			
Coding sequences below are predicted from manually evaluated			
computer analysis, using similarity information and the programs;			
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and			
GeneMark; http://opal.biology.gatech.edu/GeneMark/			

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES

source

```

1. .20753
  /organism="Salmonella typhimurium LT2"
  /mol_type="genomic DNA"
  /strain="LT2; SGSC 1412; ATCC 700720"
  /db_xref="ATCC:700720"
  /db_xref="taxon:99287"
  /note="LT2"
  complement(71. .493)
  /gene="flgN"
  /note="synonym: STM1171"
  complement(71. .493)
  /genes="flgN"
  /notes="believed to be export chaparone for FlgK and FlgJ; flagella synthesis protein FLGN. (SW:FLGN_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20101.1"
  /db_xref="GI:16419688"
  /translation="YTRLSEILDQMTVLNDLKTVDABEQQLSVQINGSLQRLITE
  EKSLIADLYEQRLRQNAQRGANDDIARWQAITKTOHLRLDNLQHNGLLEGG
  IERNQALVLPKPHQPTLYGADGQTSVSHRGKKISI"
  complement(498. .803)
  /genes="flgM"
  /note="synonym: STM1172"
  complement(498. .791)
  /genes="flgM"
  /notes="anti-sigma; negative regulator of flagellin
  synthesis (anti-sigma-28 factor). (SW:FLGM_SALTY); RflB
  protein"
  /codon_start=1
  /transl_table=11
  /product="anti-FlaA factor"
  /protein_id="AAL20102.1"
  /db_xref="GI:16419689"
  /translation="MSIDRTSPLKPVSTVOTRETSTPVOKTROKETSAAATSASVTL
  SDAQALMQQVSDINNRVEALKTAIRNGELKMDIKIADSLIREAQSLQSK"
  complement(798. .803)
  /genes="flgM"
  /notes="putative RBS for flgM; RegulonDB:STMS1H001438"
  complement(883. .1556)
  /genes="flgA"
  /note="synonym: STM1173"
  complement(883. .1542)
  /genes="flgA"
  /notes="assembly of basal-body periplasmic P ring; flagella
  basal body P-ring formation protein FLGA precursor.
  (SW:FLGA_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20103.1"
  /db_xref="GI:16419690"
  /translation="MQTLKRGFAVALLFSPUTMAQDINAQLTWFSQRLAGFSDV
  VTLRSPNLLPSCEQAFSGAKLGNVNVVAVCANEKRYLQNVQATGNYAVAA

```

```

PIARGGKLTPTANVTLKRGRLDQLPPTVLTDIRIQDAVSLRDLAPGQPVOLTMIRQAW
RVKAGQVQVIANGEGFVNAEQAMNAAVAQNRVMTSGQIVSGIVSDGNNILN
L"
complement(1551. .1556)
  /gene="flgA"
  /note="putative RBS for flgA; RegulonDB:STMS1H001439"
  1691. .2115
  /gene="flgB"
  /note="synonym: STM1174"
  1691. .1696
  /gene="flgB"
  /note="putative RBS for flgB; RegulonDB:STMS1H001440"
  1699. .2115
  /gene="flgB"
  /note="cell-proximal portion of basal-body rod; flagellar
  basal-body rod protein FLGB (putative proximal rod
  protein). (SW:FLGB_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20104.1"
  /db_xref="GI:16419691"
  /translation="MLDRDLAALRFQOEALNLRQOEILAAIANADTFGYQARDID
  FASELKKVMYRGREETGGVALTTSSHIFPAQVSSPAVDLLYRVDPQPSLDGNTVDM
  DRERTQPADNSLKYQMGTLVLGSQLKGMNVLQGN"
  2105. .2523
  /gene="flgC"
  /note="synonym: STM1175"
  2105. .2110
  /gene="flgC"
  /note="putative RBS for flgC; RegulonDB:STMS1H001441"
  2119. .2523
  /gene="flgC"
  /note="cell-proximal portion of basal-body rod; flagellar
  basal-body rod protein FLGC (putative proximal rod
  protein). (SW:FLGC_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20105.1"
  /db_xref="GI:16419692"
  /translation="MALNIFDIAGSALAQSKELNVAASNLANADSVTPGDPQPYRA
  KQVVFDAAPGQATGQKVVASVIESQAPKLYVEFGNPLADANGYKMPNVDVGM
  VNTMSRSYQANIEVLNTVKSMLKLTLLGQ"
  2523. .3233
  /gene="flgD"
  /note="synonym: STM1176"
  2523. .2528
  /gene="flgD"
  /note="putative RBS for flgD; RegulonDB:STMS1H001442"
  2535. .3233
  /gene="flgD"
  /note="initiation of hook assembly; basal-body rod
  modification protein FLGD. (SW:FLGD_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20106.1"
  /db_xref="GI:16419693"
  /translation="MSIAVNMNDFNTGKVTGTTGSGMTSGNSAADLQSSFTLLVAQL
  KNQDTPNQLNNELTQIAQISTVSGIKLNTIGALSQIDNSQSQAATLIGHGYM
  VPGTTLAGKABEGAVTSTPFGVELQQPADKVTAITDKDGVVVTLEIGELRAGV
  HTFTWDGKQDGTTPNGSYNIALTSANGCTQLVAQLQALVQGVTKSGNGLDLG
  TYGTTLDDEVNQI1"
  3248. .4471
  /gene="flgE"
  /note="synonym: STM1177"
  3248. .3253
  /gene="flgE"
  /note="putative RBS for flgE; RegulonDB:STMS1H001443"
  3260. .4471
  /gene="flgE"
  /note="flagellar biosynthesis; flagellar hook protein

```


REFERENCE AUTHORS	Enterobacteriaceae; Salmonella. 1 (bases 1 to 254050) Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leatherbarrow, S., Moule, S., O'Garra, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G.
TITLE	Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18
JOURNAL	Nature 413 (6858), 848-852 (2001)
MEDLINE	21534947
PUBMED	11677608
REFERENCE AUTHORS	2 (bases 1 to 254050) Parkhill, J.
TITLE	Direct Submission
JOURNAL	Submitted (25-Oct-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT	E-mail: parkhill@sanger.ac.uk
NOTES	Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES	Location/Qualifiers 1..254050 /organism="Salmonella enterica subsp. enterica serovar Typhi" /mol_type="genomic DNA" /strain="CT18" /db_xref="taxon:90370" 181..2793 /gene="STY1078" /note="synonym: pepN" 181..2793 /gene="STY1078" /note="Orthologue of E. coli pepN (AMPN_ECOLI); Pasta hit to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
gene	/codon_start=1 /transl_table=11 /product="aminopeptidase N" /protein_id="CAD08183.1" /db_xref="GI:16502232" /db_xref="GOA:Q8Z7T0" /db_xref="TrEMBL:Q8Z7T0" /translation="MTQQPQAKYHRYADPYQITDIETLFDLDAAKTVVTAISOAVR HGAPDARLGDGTLVSIHVNDAPWATYKEEGALLISDLPERFLRIYNIESPA RNTALGELVQSDALCTQCEAGFRHTWYLDPRDVLTARFTTKIAIDKSPFLLNSGN RVAQGELENRWTHVQWQDFPKPKYFALVAGDEVDLRDTFTTRSGREVALEYLVDRG NLDRAPMANGTSIKSMKWMDQRFGLEVDLDTIMIVAVDFNNMGAMENKGILFNISKYV LARTATDKYLDITERVIGHEYFHNTGNRVCRDWFSQLSKGLGVTPRDQSFSDL GSRAVNLNVTFMRLOPADASPMAPHIPDPKVIEANNFYTLTVVEKAVERMIH TLIGENFOKGMQLYFERHDSGAATCDDFVQAMEDASNVLSHFRWYSQSQTPIVTV KDNYPTAEQTYLTISRTPTATAQAEPQHIFPALEIXDNGSNVIPFOKHGPVNA VINDVTAQETFTFDNFYQPVPALLCEFSAPVKLEYKWSQQLTFLMHRNDFSWD AAQSILATIKLINVARHQGGPFLSLPVHADAFPAVLLEDKIDFPALAAEITLPSANE IAEPLFVIDPIAIAQVRALNRITLAELADEFIAIYNANHLDREVYDHGDIKRTLNR ACRLFTWFTEILTANLSKYOVDRANNMTDALASAVALPCPDRTMOEYDDKWH QOGLVMDCFEELIQTSAPSVNLETVSRLIKHRESMSNPRIEISLCAFAGSNPAFH AQGGGYQLVEMLDLNSRPQVASRLIEPLIRLKRYDDKRQEKKRAALEQLKGLEH LSGDLYEKIKTALA" 532..1332 /gene="STY1078" /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase family M1, score 245.10, E-value 1.9e-84"
misc_feature	1060..1089 /gene="STY1078" /note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature" 3001..4011 /gene="STY1079"
gene	

CDS	/note="synonym: pyrD" 3001..4011 /genes="STY1079" /notes="Orthologue of E. coli pyrD (PYRD_ECOLI); Fasta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap" /codon_start=1 /transl_table=11 /product="dihydroorotate dehydrogenase" /protein_id="CAD08184.1" /db_xref="GI:16502233" /db_xref="GOA:Q8Z7S9" /db_xref="Swiss-Prot:Q8Z7S9" /translation="MYYPVRKALFQDLPERRAHEFTFOOLRRITGTPLEALVROKVP KPVTOMGLATFKNPLGACLGDKGCIDALGAMGFSGLEIGTVTTPROPGNDDKRLPR LVDAEGLINRMGFNGLVDNLVENYKKAHFQGLIGINTGKNKDTPENGKNDYLICME KYAYAGYIAINISSPNTPGLRTLIQYGADLDLLTAIKNKQNDLQVIHKHYVPYAVKI APDLCEEILQVADSLRNIDGVIATNTILDRSLVQGMKNCQQTGGSUGSRPLQKST EIIRRLSELKGQLPTIIGVGSDISVIAAREKITAAGATLVQIYSGFIKPGPLIKEIV HI"
misc_feature	3136..3990 /gene="STY1079" /note="Pfam match to entry PF01180 DHodehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165" 3238..3297 /genes="STY1079" /notes="PS00911 Dihydroorotate dehydrogenase signature 1"
misc_feature	3874..3936 /gene="STY1079" /notes="PS00912 Dihydroorotate dehydrogenase signature 2"
gene	4174..4719 /genes="STY1080" 4174..4719 /genes="STY1080" /note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
CDS	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAD08185.1" /db_xref="GI:16502234" /db_xref="TrEMBL:Q8Z7S8" /translation="MMRIKPPDNWRYYDEEHDRWMDLANGMLFRSFRSKMLTPDA FCPTGFCVDDAALYFSEKPKDRDFELTKEQRAELVNALVAIRLYLKPMQKSWHFVAH GENWTPTGTGDAASVWLSDTAEQVNLVVPEGENAALCLLAQPGVVVIAGRTMQLGDAIK IMNDRLLKPVOCHCSFSLEQAV" complement(4716..5825) /genes="STY1081" complement(4716..5825) /genes="STY1081" /notes="Orthologue of E. coli P75863; Fasta hit to P75863 (369 aa), 86% identity in 367 aa overlap" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAD08186.1" /db_xref="GI:16502235" /db_xref="GOA:Q8Z7S7" /db_xref="TrEMBL:Q8Z7S7" /translations="MATLSRLTFHPVKSMRGIGLTHALDISGLADFIFMVTESDGT PITATARPQMVRTPSPDLGHGLTAPDGSSALVFRTDTPDQAPTEVWGHNHRTARVA PTAINQWLSGFFRSRDVQLCWGSQLTRERKXHNVAPLGFADGPVYLLNEASRLDLQ PCPAGVOMEQFRNLVYSGVAAWEEDSKVYLRIGDIVIDVVKPCSRCIFTTVSPEKQG KHFSGBPLTLQAFRTAQDNGVDVFGQNLIARSNVIRGVGEVEILATAPAKAGAIT LDPSVTPRKHPDGSVTDHWQGTFCGNQVQLLEGQENQVIRIPYSCRAGICGCCGRIR LLSEGEVSPKKSAMGDDGITLSCSCVPTALREN" complement(4737..4955) /gene="STY1081" /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13" 5924..8032
misc_feature	
gene	

gene	RAINALGATPMDLMSILQNSQAGCLRAKLEII"	Score: 498.00	Matches: 144
	/complement (5597. .6295)	Percent Similarity: 43.91%	Conservative: 76
	/genes="figH"	Best Local Similarity: 28.74%	Mismatches: 167
CDS	/locus tag="t1740"	Query Match: 19.26%	Indels: 114
	/complement (5597. .6295)	DB: 1	Gaps: 14
	/genes="figH"	US-10-009-823A-1 (1-502) x AE016840 (1-301983)	
	/locus tag="t1740"	QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29	
	/notes="corresponds to STY1219 from Accession AL513382:	Db 9112 AGCGGCTGAACGCTGGCGCCACCACTTGATGTTATCGGTAATACATCGCACTCC 9053	
	Salmonella typhi Crl8"	QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49	
	/codon start=1	Db 9052 GCCACCTATGCTTAAAGTCCGGTACGGCATCATTTGCCGATATGTTGGCC----- 9002	
	/transl table=11	QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69	
	/product="flagellar L-ring protein precursor"	Db 9001 -----GGTTCC-----AAAGTGGGTTGGCGGTAAAGTGGCG 8969	
	/protein_id="AA069364.1"	QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89	
	/db_xref="GI:29137803"	Db 8968 GGGATTACCCAGGATTTTACCGAGGTACGACAAACACCGCGCGCGGCTGGATGC 8909	
	/translations="MKYALHAYVVMALMVAITLGCWAPKPLVOGATTAQIPGPV	QY 90 AlaIleGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108	
	PVANGIFQSAQINTGYQPLFEDRPRIQDITLVLQENVSASXSSANASRDGKT	Db 8908 GCGATTAGCCAGAACGGTTTTTTCGCTGTGTAGACAGCAACGGTTCGTGTTCTATAGC 8849	
	SFGDTPVRYLOGLFGNSRADMGSGNSFNGKGANASNTFSGLTITVTDQVLANGN	QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128	
	LHVVGKQIAINGTGFIRPFGVNVNPTTISGNSVPSTQVADARIEYVGVNGYNEAQN	Db 8848 CGCAACCGCCAGTTCAAACGAGACGAGAACCGTAACCTGTCAATGTCAGGGGATGCAG 8789	
	XGMLQRFNLNLSM"	QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148	
	/complement (6350. .7132)	Db 8788 TTGACCGGCTATCGGCCACCGGTACCGCCGACCATTCAGCAGGGGGCGGAATCCTCGC 8729	
	/genes="figg"	QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168	
	/locus tag="t1741"	Db 8728 CCGATCACCATTCGGAACACGCTGATGGCG-----GCGAAATCGACCAACCGCGC 8678	
	/notes="distal rod protein; corresponds to STY1218 from	QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188	
	Accession AL513382; Salmonella typhi Crl8"	Db 8677 TCAATGACAGATCAACCTG---AACTCAACGACCTGTACCGTCTAAACG----- 8630	
	/codon start=1	QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSer----- 206	
	/transl table=11	Db 8629 -----CCCTTTAGCGTGGATGCG 8609	
	/product="putative flagellar basal-body rod protein FlgP"	QY 207 ---AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225	
	/protein_id="AA069366.1"	Db 8608 GATTCGTATACAAAAAGGACCGTCACCTGTTTATGACAGCAGGGTATGTCATGAC 8549	
	/db_xref="GI:29137805"	QY 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245	
	/translations="MDHAIYTAMGAASQTINQAVTASLANASTPGFRAQLNALRAV	Db 8548 ATGAACGCTCTATTT-----GTGAAAACCAAGATAATAATGAATGGCT 8507	
	PVDELATLTATSTPGADMTQGLDYTSRPLDVALQDDGMLVVAADGAEGYTRN	QY 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265	
	GNVQPTGTLTQGHPIVIGEGPIVTPGSEITIAADGTISALNPGDPNTVAVPGR	Db 8506 GTGTACACCCATGACAGCAGCATCTGTCAGCCACTCGCCCAACACGCGG----- 8456	
	LKLVKAEENSVQSDGLPRLTAEQAERGAVLAADPSIRINSGVLGNSNKPVEAMT	QY 266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys----- 280	
	DMANARRPMQMKVITSVDENEGRANQLLSMS"	Db 8455 -----TCCACTACGCTGAATTCATGAAAACGGGATTCGGAGTCTGGCGGTACGGTG 8402	
	/complement (7922. .9133)	QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300	
	/genes="figE"	Db 8401 AACATCACCCCGGTACGATTAATGGCGCACAGC----- 8366	
	/locus tag="t1743"	QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlnIleGln 320	
	/notes="corresponds to STY1216 from Accession AL513382:	Db 8365 -----GCCACCTTCTCCCTCAGCTTCTCTTAATCC----- 8336	
	Salmonella typhi Crl8"	QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340	
	/codon start=1		
	/transl table=11		
	/product="flagellar hook protein FlgH"		
	/protein_id="AA069367.1"		
	/db_xref="GI:29137806"		
	/translations="VSFSAVSGCLNAAATNLVDIGNNIANSATYGFKSGTASFADMEA		
	/db_xref="GI:29137806"		
	/translations="VSFSAVSGCLNAAATNLVDIGNNIANSATYGFKSGTASFADMEA		
	GSKVGLGVKAVGATQDFTDGTNTTNGRLDVALSQNGFFFLVDNSGVSFYSRNGQFKL		

Alignment Scores:
Pred. No.: 1.22e-20 Length: 301983


```
/locus tag="c1256"
/notes="Residues 18 to 169 are 94.07 pct identical
to residues 1 to 152 of 152 from SwissProt.40 :
>sp|Q57334|T200_SALTU Transposase for insertion sequence
element 18200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79730.1"
/db xref="GI:26107531"
/transl_table=11
/translation="MSGCKKQTRNQKVPNGNKSIAHTRWCKYHIVFAPKVRQ
VFYKRRRAIGCILKLEWKSRIIEACCCADHIMLVEIPPMVSFGFMGLKGS
SLMPYQFQGLKFKYRNRFWCRCGYVDVTKGNKTAQIDYIKHQLEEDKMGQLSTPY
PGSPFTGRK"
/complement(6578..6940)
/notes="incomplete"
/insertion seq="IS4"
/complement(6598..6960)
/locus tag="c1257"
/complement(6598..6960)
/locus tag="c1257"
/function="IS, phage, Tn"
/notes="Residues 8 to 120 of 120 are 98.23 pct identical to
residues 316 to 428 of 442 from MG1655 : b4278"
/codon_start=1
/transl_table=11
/product="Putative conserved protein"
/protein_id="AAN79731.1"
/db xref="GI:26107532"
/transl_table=11
/translation="MPROPDQYREIKOTMOLSRITLRSKPKBELVBOELWGLVAYNL
VYQMLKMAHSLAGYVNPOLSPFESCGWYVMTLQASPGRIPELMDRLASMGOLV
KLPTRETAFFRVKERP"
/locus tag="c1258"
/locus tag="c1258"
/locus tag="c1258"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79732.1"
/db xref="GI:26107533"
/transl_table=11
/translation="MLGACFYSGMTGYFGFPMWGAELVMMWPEACFLLQSSLYSL
LDQVLFPSFWELNKQYEGFEFYQIFSGNIVASCSSGVSNSAHYARCTDFWRH
YSLNNVAAACOLGRQACLS"
/locus tag="c1259"
/locus tag="c1259"
/locus tag="c1259"
/notes="Residues 1 to 139 of 144 are 50.71 pct identical to
residues 104 to 242 of 258 from GenPept.129 :
>gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
[Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79733.1"
/db xref="GI:26107534"
/transl_table=11
/translation="MKDLSFIRFFLAILLVPPALFFVHRHITPGALTLCVITAL
IITLLASGLRYSVKCISELSIFRHRIECHMAGPMIYFTFSLIAGWKYKQFKKEY
QTLHYNARYIVLARDGLRLVLSQYSRSRFRVIINGHRL"
/locus tag="c5534"
/locus tag="c5534"
/locus tag="c5534"
/product="tRNA-Arg"
/function="tRNA-Arg"
/complement(8832..9290)
/locus tag="c1260"
/complement(8832..9290)
/function="putative enzyme; Transposases Insertion
Sequence Associated"
/notes="Escherichia coli O157:H7 ortholog: z4324"
```

```
/codon_start=1
/transl_table=11
/product="Putative Transposase"
/protein_id="AAN79734.1"
/db xref="GI:26107535"
/translation="MASAADARRIVSHYERRNLIEEVHKAKWGGTCTVESLRMQTRDN
LERMVVIAFAVAMLGLRQEGISEETONDSCKILPTTEWKLWVLEKGLQSPQP
TLKWACLKGRWHDSTKRTGPGWVWMDWFRLODMVEGYFVAKSLDQET"
/complement(9302..9667)
/locus tag="c1261"
/complement(9302..9667)
/locus tag="c1261"
/notes="Residues 1 to 35 of 121 are 97.14 pct identical to
residues 59 to 93 of 316 from GenPept.129 :
>gb|AAK18437.1|AF348706_126 (AF348706) putative
transposase [Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79735.1"
/db xref="GI:26107536"
/translation="MABRLGEIQKRVITCDREADIWHYLYHKVSHGQREVFVTAQNS"
Alignment Scores:
Pred. No.: 1-24e-20 Length: 306358
Score: 498.00 Matches: 143
Percent Similarity: 43.57% Conservative: 74
Best Local Similarity: 28.71% Mismatches: 171
Query Match: 19.26% Indels: 110
DB: 1 Gaps: 15
US-10-009-823A-1 (1-502) x AE016759 (1-306358)
QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 73424 AGCGGATTAAACGGTGCCTGCCACCACTCGATGTTATTGGCAACAATATCGCACTCC 73483
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 73484 GCCACCTACGGCTTTAAATCAGGCACGGCTCTTTTCCGATANGTTTGGC----- 73534
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 73535 -----GGTTTCG-----AAAGTGGGACTCGGGGTAAAGTTGCT 73567
QY 70 SerValargThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 73568 GGTATCACTCAGGACTTTACCGATGCGCAGCACCAACACCGGGCGTGTCTGGACGTT 73627
QY 90 AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108
Db 73628 GCTATCAGCCAGAACGGTTTTTTCGCTGTGTAGACGACCAACGGTTCGGTATTCTACAGC 73687
QY 109 ArgAlaGlyAsnPheAspPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 73688 CGTAACGGACAATTAAAGCTGGATGGAACCGTAACCTGTGAATATGCAAGGTTTACAG 73747
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 73748 CTGACGGGTACCGGCAACCGGTACCGCCGCGACTATTACGACAGCGGGCGAATCCGACC 73807
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 73808 AATATTTTCGATCCCGAATACCTCGATGCGCA-----CGGAAACTACCAACGCGCG 73858
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 73859 TCGATGAGATCAACCTG---AATTCAGTATCGGTTCCCTCTGTT----- 73903
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr 208
Db 73904 -----AACGCATTTGATCCGCAATGCGGATGCTAT 73936
```

```

QY 209 SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrVal 228
Db 73937 ACAACAAAAGGTTGGTCTGCTGTTTCAGTATGACGGGTAATGCTCATGATGACGCTG 73996
QY 229 TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlyTyrLeuValAlaMet 248
Db 73997 TACTTTT-----GTGAAGACCGGGGATATAAGTGGCAGCTCTAC----- 74035
QY 249 AsnProSerGluAspGlySerAlaLaserGlyThrAspSerAlaGlyLeuLeuMetSer 268
Db 74036 -----ACCCAGGATAGCAGTGCACACAGGTACAGCTGAGCTGCAATG----- 74080
QY 269 GlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThr 288
Db 74081 ---ACCTCGTGTGTTAATCCCAATGCGTCTCG-----ACCTGCTTAC----- 74113
QY 289 GlySerAlaThrLysAsp-----LeuAsnAlaTrpGlnProAlaProLeu 303
Db 74114 TCAGATCCAAACAAAGGATATTACCCAGCGCAATTAACGGCGCAGATCCGCC----- 74167
QY 304 ValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThr 323
Db 74168 -----ACGTTTACCTGAGCTTCTCCTCAACTCC----- 74194
QY 324 LeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 343
Db 74195 -----ATGCAGCAAAAT----- 74206
QY 344 AlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGly 363
Db 74207 -----ACCGCGCTTAAC 74218
QY 364 AsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrPro 383
Db 74219 ALCATTGTGGCAACC-----ACCGAACCGGCTTACAAA 74251
QY 384 GlnGlyAspLeuValAspValThrIleThrSerGlnGlyLysLeuGlnGlyLysTyrSer 403
Db 74252 CCGGGCGATCTGCTGATTCATCAATGATGACGGTACGGTGTGCGCACTATTCC 74311
QY 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
Db 74312 AACGAACAAACCAACCTGCTGGCGGAGATGTACTGGGCACTTTTCCCAACACGAGGT 74371
QY 424 LeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Db 74372 CTGGCATCGAAGCGGACACCTGCTGGTCCGCTACGCAATCTTCTGGCGTGGCGCTGTG 74431
QY 444 GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463
Db 74432 GGGACACCGCGGACCGGCAACTTGGCACACTTGACACACTGACCAACGGTGGTGAACGCTCCAAC 74491
QY 464 ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483
Db 74492 GTTCATCTCAGTAAGAACTGTGTCATATGATGCTGTGCCACGGTAACATCATGCTTAAC 74551
QY 484 SerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 74552 GCCAGACCATCAAAACCCAGGACCAAGATCTCTCAACACGCTGGTTAACTTACGC 74605

RESULT 28
AE005317
LOCUS 9224 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 141
of 155.
ACCESSION AE005317
VERSION AE005317.1 GI:12514603
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 9224)

```

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.

Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 Nature 409 (6819), 529-533 (2001)

2 (bases 1 to 9224)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.

Direct Submission

Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

Location/Qualifiers

1. 9224

/organism="Escherichia coli O157:H7 EDL933"

/mol_type="genomic DNA"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohaemorrhagic"

134. 718

/gene="rimJ"

/note="synonym: Z1703"

134. 718

/gene="rimJ"

/functions="enzyme: Macromolecule synthesis, modification: Ribosomes - maturation and modification"

/notes="Residues 1 to 194 of 194 are 100.00 pct identical to residues 1 to 194 of 194 from Escherichia coli K-12. Strain MG1655: B1066"

/codon_start=1

/transl_table=11

/product="acetylation of N-terminal alanine of 30S ribosomal subunit protein S5"

/protein_id="AAG55812.1"

/db_xref="GI:12514604"

/translations="MFGYRSNVKPKVLTDLRVLRVLDHRAWLADYYAENRHLKP WEPVRBHCYQSGQWQARLQMEFHQKSAFYGLFDPDEKEIIIGVANSNVVRSF HACVLGYSIGKQKQGLMFEALTAIRYMORTQIHRIWMYMPHMKRSGDLLARLG PEKGYAKDYLIDGGWRDHLVLTALTPDWTGCR"

729. 1376

/gene="yceH"

/note="synonym: Z1704"

729. 1376

/gene="yceH"

/functions="orf; Unknown function"

/notes="Residues 1 to 215 of 215 are 99.06 pct identical to residues 1 to 215 of 215 from Escherichia coli K-12 Strain MG1655: B1067"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAG55813.1"

/db_xref="GI:12514605"

/translations="MKYQLTALAEARVIGCLLEKQVTPPEQYPLSVNGVVTACNQKTR EPWNVSESVQQLNIVRRHYLRVSGFNGNRVTKYEQFCNSEFGDLKLAAEVAL IYTLRLGACQTPGLSRRAARMYFEDMAEVELTQLANREDGPFVRLARPGKRE SRYMHLFSGSEVEDQPAVTDMNSAVDGLQARVEALEIAVELKQRLDLSLAHLGD"

1378. 2301

/gene="mvim"

/note="synonym: Z1705"

1378. 2301

/gene="mvim"

/functions="putative factor; Not classified"

/note="Residues 1 to 307 of 307 are 99.02 pct identical to

Ohnishi, Keisuke Nakayama (Miyazaki Medical College), Kozo Makino, Ken Kurokawa, Katsushi Yokoyama, Masashi Tanaka, Takashi Honda, Teruo Yasunaga, Hideo Shinagawa (Osaka University), Takahiro Murata (Shinshu University), Chang-Gyun Han, Eiichi Ohtsubo, Toru Tobe, Chihiro Sasakawa (University of Tokyo), Hideto Takami (Japan Marine Science and Technology Center), Naotake Ogasawara (Nara Institute of Science and Technology), Satoru Kuhara (Kyushu University), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science.

FEATURES

source

1. 222605
Location/Qualifiers
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
74. 607
/gene="ECs1423"
74. 607

/gene="ECs1423"
/note="similar to B1045_ECOLI gi|1787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative polyprotein"
/protein_id="BAB34846.1"
/db_xref="GI:13360887"
/translation="MKTRHVGQDITKLVAVVNVNANPSLMGGGVGATHRAAGP
ALLDACLVQQQDCPTGHAVITLAGDLPKAVHTVGVWGEQNDQLLQDLYL
NSLRNVAANSYTSVAFAISTGVIGYPRAAASIAVKTVSSEFTRHALPEQYIFVCYD
EENAHYERLLTQGDSE"
549. 2030
/gene="ECs1424"
549. 2030

gene

CDS

/gene="ECs1424"
/note="similar to YMC_ECOLI gi|1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative synthase"
/protein_id="BAB34847.1"
/db_xref="GI:13360888"
/translation="MMKKTPTSKDLPNKEMNDLPRLASAVLPICSPQPGCGLPPL
EKSLDPAARLRAEMSEHTLDVQYIWDMSGRLLFSAALAAKRGVVRLLDDN
NPLGDLIALLSHPRIEVLFNPFSLRLLPLGYITDFSLNRMRNKSFTVGVV
TLVGRNIGDAYFCAGEEPLFSLDVAIGPVVEDVADDFARYWCKSVPLQQLVDV
PEGEMADRIELPASWINDAMTHRYLRMESSPFILHVDGTPLIWAKTLLSDDPK
GEGKAKHSLPDLFDINGSERDIISSVEFVTRAGVALLRWKGVKIAIITN
SLANDVAVHAGYAKRWKLLRYGVLELHELKTRQOSTLDHGTGNSGSLHAKT
FSIDGKTVFGSFPNRTLLTMEGFVIESETLAQLDKRPIQSCYDAQAQLRLDR
WGRINWDRAKKEIIVLKKEPASFVKRWVRLASILPVEWLL"
complement(2038. 3195)
/gene="ECs1425"
complement(2038. 3195)
/gene="ECs1425"
/note="similar to MDOC_ECOLI gi|3025226sp|P75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="glucans biosynthesis protein"
/protein_id="BAB34848.1"
/db_xref="GI:13360889"

gene

CDS

/translation="MNPVPAQRYFDISRAWMLLGIPIHSLIYSSTWHVNGAEP
SLMLTFDHFIRVQVFFVIGSYFMYLFLRYPLKWKWVRVGRVIMPLTAPLL
TLQFIMLVQKGAWSVGLSYDKNTLAWELISHLWFLVLVMTTLCVMTFKRI
RNNLSDKTKNKVSMVKSYFELCIGIVAVIRRTFIVYPPILSGNFVIMQTL
FYLPFTLALATIFPHLKALFTTPSRGCTLAALAFVAYLLNORYSGDAMWYTES
VTWVLGLWNVVFSFGRLLNFQSAARTVFNASLIYLVHHPITLFFGAYITPHI
TSWNLGFLCGLI FVVGITAILYBIHLRIPILKFLFGKPVKRENDKAPAR"
3588. 5123

gene

/gene="ECs1426"
3588. 5123
/gene="ECs1426"
/note="similar to MDOC_ECOLI gi|1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translation="MMKORWLSAAVMTLTSSSWAFSIDVAKQASLACKGYEAPK
SNLPSVFRDMKYADYQIQFNHDKAYNNLTKTFKLEFYHCGMYFDTFVKINEVTATA
VRIKYSDFYTFDGVQDKDVTDLGDFAGFKVLYIPINSKDNDEIVSMLGASVFRVI
GAGOVYLSARGLAIDTALPGSEFPFRKFWIERPKPTDKRLTIYALLDSPRATGAY
KFVMPGRDVTVDVQSKIYLRDKVGLGVAPLTSMPLFGNPQSPANNYPERLHDNG
LSIHAGGEMIRPLNPKHLAVSSPSMENPQGLLQRCGRDFSRDELDLDRVLDLPS
AWPTPKGKMGKSGVELVEIPTNDETNDNIYAVTTPOLPEPKEMKFKTITTSRDED
KLHAPDNAAVQQRTRSDGVQNLIRKQPDGTIAFVVDFTGAEMKKLPEDTPTAOTS
IGDNGEIVESTRYNPVTGKMRVMRVKVDKAKKTTEMAALVNADQTLSETWSYQLP
ANE"
5086. 7659
/gene="ECs1427"
5086. 7659
/gene="ECs1427"
/note="similar to membrane glycosyltransferase MDOH_ECOLI
gi|1787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MKPGATSYLPNMKNTTYIDAMPFAEKAALPKTDIRAVHQAID
AEHTWAREDSDPGSVKARLEQAPSDLAGQLIKDDGRDOLKAMPEAKRSMPFD
PWRTNPGVFRDRLGRDTPRYLARTKEOSEQKRWTVGTIRRVILILTAQTV
VATYMKITILPYQGHALLNPMQVGDLYWSEWQLLYMLQGTGILLFVLFQWSAG
FWTALMGFLQLLIGRDKISASTVGDPEPLNPHRTALINPI CNEDVNRVFAGLRATW
ESVATGNKAKHFDVYILSDSYNDPICVAEQKAWELIAEVGGGQIFRRRRRVKPK
SGNIDDFCRWGSQSYVMVLDADSVNTDCLCLVRLMEANFAGLIQSSKASGMD
TLVARCOQFATRVYGPITAGLFWQGESYTHGNAIIRVKKPFIHCALAPLPGES
FAGSILSHDFVEAALMRAGVWIAVYDLPGSVEELPNLLDELKDRWCHGNLKNF
RLVFGMHPVHRAVFLTGVMSYIAPLWFMFLALSTALQVHALTEPOYFLOPROLF
PWWQWPELALFASMTVLLFLPKLLSILLNCKTKYGGFWVTLISLLEVLFS
VLLAPVRLMFLAPVFSLLSPSVISRAATVGLTKRWKFLIPIEESYSPQVLVD
MAWLDLFLFWLAPIVFSLLSPSVISRAATVGLTKRWKFLIPIEESYSPQVLVD
TDRELNRRORSDDGFHAFVNPSPNALATAMARHASKVLEIARDHRVBOALNE
TPEKLNDRRLVLLSDPVTMARLHFRVMSPERYSWVSYYEGIKLNLPLALRKPDAAS
Q"
7832. 8059
/gene="ECs1428"
7832. 8059
/gene="ECs1428"
/note="identical to YCEK_ECOLI gi|1787288 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MRLIVVSIWVTLSSGGGSIISRTIPQGGHGNQYYPGVQWMDVRDS
AWRVYTLIDLPFSLVFDLTLPLIDIHGPEY"

Alignment Scores:

Pred. No.:	1.8e-20	222605
Score:	493.00	142
Percent Similarity:	44.13%	76
Best Local Similarity:	28.74%	174
Query Match:	19.06%	102
DB:	1	15

QY	10	ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla	29
Db	26851	AGCGGATTAACACGCTGCCCAACCACTCGATGTTATGGCAACAATATCGCAACTCC	26910
QY	30	AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu	49
Db	26911	GCCACCTACGGCTTTAAATCAGCAGCGGCTCTTTTGGCATATGTTTGGC-----	26961
QY	50	AlaIleGlySerThrClySerGlnGlyProAsnGlnAlaGlyMetClyAlaGlnValCly	69
Db	26962	-----GGTTCCG-----AAAGTGGGACTGGGGTAAAGTTGCC	26994
QY	70	SerValargThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu	89
Db	26995	GGTATCACTCAGGACTTTACCGATGCGACGACCAACACCGCGCGTGTCTGCAGCTT	27054
QY	90	AlaIleGlyClyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr	108
Db	27055	GCTATCAGCCAGAACGGTTTTTTCGTCGTGTAGATAGCAACGGTTCGGTGTCTTACAG	27114
QY	109	ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr	128
Db	27115	CGTAACGGCAATTTAAAGCTGGATGAAATCGTAACCTGGTGAATATCAAGGTTTACAG	27174
QY	129	LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle	148
Db	27175	CTGACGGGTACCCGCCAACCGGTACGCCGCACTATTTCACAGAGGGCGGAATCCGACT	27234
QY	149	GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
Db	27235	AACATTTTCGATCCGAATACCTCGATGGCA-----GCGAAACTACCAACACGGCG	27285
QY	169	AsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
Db	27286	TCGATCGCACTCAACTCG---AATTCCAGCATCCGGTCCCTCTCTGTT-----	27330
QY	189	PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr	208
Db	27331	-----AACGCATTTGATGCCCAACTGCGGATGAGCTAT	27363
QY	209	SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisaspIleThrVal	228
Db	27364	AACAAAAAGGTTCCGTGACTGTTTCGACAGTCAGGGTAATGCTCATGACATGAGCTC	27423
QY	229	TyrPheAspGlyAlaProSerSerThrThrGlySerLysThrPheGluTyrLeuValAlaMet	248
Db	27424	TACTTT-----GTGAAGACCGGGGATAAATCACTGGCAGCTCTAC-----	27462
QY	249	AsnProSerGlnAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer	268
Db	27463	-----ACCCAGATGACGTGATCCACAGGTACAGCGACGCTCGCAATG-----	27507
QY	269	GlyThrMetThrPheSerSerAsnGlyGluLeu---LysAsnMetThrAlaPheThrPro	287
Db	27508	---AAGCTGGTGTTTAATGCCAATGCGTCTTGACCTCAATCCACAGAGAATATTACC	27564
QY	288	ThrGlySerAlaThrLysAspLeuAsnAlaTTrpGlnProAlaProLeuValAsnGlyLeu	307
Db	27565	ACCGCGCA-----ATTACGGCGGAGAACCCGCC-----	27594
QY	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly	327
Db	27595	---ACGTTTAGCTGACTTCTCTCAACTCC-----	27621
QY	328	IleLysSerGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr	347
Db	27622	-----ATGCAGCAAAAT-----	27633
QY	348	AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla	367
Db	27634	-----ACCGGCGCTTAACAACTTGTGCA-----	27657

QY	368	ArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu	387
Db	27658	ACC-----ACCAGAAATGGCTCAACAAACGGGGCATCTG	27690
QY	388	ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal	407
Db	27691	GTGAGTTATCAATCAATGATCAGGTACGGTGTGCGCAACTATTCCACGAACAAC	27750
QY	408	ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu	427
Db	27751	CAACTGCTGGGGCAGATTGTACTGGCGAACTTGCACAAACGAAGGCTCGCATCCGAA	27810
QY	428	GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly	447
Db	27811	GGCGACACAGTCTGGTCTGCGACGCAATCTTCTGGGGTGGCGTGTGGGACAGCCGG	27870
QY	448	ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer	467
Db	27871	ACGGGCACTTTGGGACCCCTGACCAACCGTGGTGGGAAGCGTCCAACGTGCTCAGT	27930
QY	468	ArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerVal	487
Db	27931	AAAGAACTGGTCAATATGATCGTTGCCACGCGTAATCATGATCTAAGCCCGACCATC	27990
QY	488	ThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys	501
Db	27991	AAAAACCCAGGACAGATCCTCAACACGCTGGTTAACTTACGC	28032

RESULT 30
 U00096 11
 WPCOMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096
 Fragment Name Begin End
 U00096_00 1 110000
 U00096_01 100001 210000
 U00096_02 200001 310000
 U00096_03 300001 410000
 U00096_04 400001 510000
 U00096_05 500001 610000
 U00096_06 600001 710000
 U00096_07 700001 810000
 U00096_08 800001 910000
 U00096_09 900001 1010000
 U00096_10 1000001 1110000
 U00096_11 1100001 1210000
 U00096_12 1200001 1310000
 U00096_13 1300001 1410000
 U00096_14 1400001 1510000
 U00096_15 1500001 1610000
 U00096_16 1600001 1710000
 U00096_17 1700001 1810000
 U00096_18 1800001 1910000
 U00096_19 1900001 2010000
 U00096_20 2000001 2110000
 U00096_21 2100001 2210000
 U00096_22 2200001 2310000
 U00096_23 2300001 2410000
 U00096_24 2400001 2510000
 U00096_25 2500001 2610000
 U00096_26 2600001 2710000
 U00096_27 2700001 2810000
 U00096_28 2800001 2910000
 U00096_29 2900001 3010000
 U00096_30 3000001 3110000
 U00096_31 3100001 3210000
 U00096_32 3200001 3310000
 U00096_33 3300001 3410000
 U00096_34 3400001 3510000
 U00096_35 3500001 3610000
 U00096_36 3600001 3710000
 U00096_37 3700001 3810000
 U00096_38 3800001 3910000

U00096_39 3900001 4010000
 U00096_40 4000001 4110000
 U00096_41 4100001 4210000
 U00096_42 4200001 4310000
 U00096_43 4300001 4410000
 U00096_44 4400001 4510000
 U00096_45 4500001 4610000
 U00096_46 4600001 4639675
 Continuation (12 of 47) of U00096 from base 1100001 (U00096 Escherichia coli K-12 MG1655)

Alignment Scores:
 Pred. No.: 9,78e-21 Length: 110000
 Score: 491.50 Matches: 143
 Percent Similarity: 43.95% Conservative: 75
 Best Local Similarity: 28.83% Mismatches: 173
 Query Match: 19.01% Indels: 105
 DB: 1 Gaps: 15

US-10-009-823A-1 (1-502) x U00096_11 (1-110000)

Qy 10 ThrGlyMetHisThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
 Db 31818 AGCGGATTAACGCTGCGCCGACCACTCGATGTTATTGGCAACATAATCGCAACTCC 31877
 Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 Db 31878 GCCACCTACGCTTTAAATACGACGCGCTCTTTGCCGATATGTTGCC----- 31928
 Qy 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
 Db 31929 -----GGTTCG-----AAAGTGGGATGCGGGGTAAAGTTGCC 31961
 Qy 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
 Db 31962 GGTATCTACTCAGGACTTTTACCGATGGCGACGACCAACCGCGGAGGCTCGGACGTT 32021
 Qy 90 AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108
 Db 32022 GCTATCAGCAGAACGCTTTTTCGCTGCTGTAGACAGCAACGCTTCGTTGTTCTACAGC 32081
 Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 32082 CGTAACGGCAATTTAAGCTGGATGAAACCGTAACCTGTGTAATATGCAAGGTTTACAG 32141
 Qy 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
 Db 32142 CTGACGGTTACCGGCAACGCGTACGCCGCCGACTATTACGCAAGGGCGAATCCGACC 32201
 Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 Db 32202 AATATTTCGATCCGCAATACCTCGATGGCA-----CGCAAACTACCAACCGGCA 32252
 Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 Db 32253 TCGATGAGATCAACCTG-----AATTCAGTGATCCGCTTCCTACT----- 32294
 Qy 189 PheAlaLeuLeuGluSerThrLysGlyAsnGlyThrProPheLysSerThrSerAsn--- 207
 Db 32295 -----GTTAGGCCATTCAGCGCCGCAATGCG 32321
 Qy 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnLysAsnSerHisAsp 225
 Db 32322 GATAGCTATAACAAAAGGTTCCGATGCTTTTCGACAGTCAAGGTAATGCTCATGAC 32381
 Qy 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
 Db 32382 ATGAGCTCTACTTT-----GTGAAGACCGGGGATAAATACTGGCAG----- 32423
 Qy 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
 Db 32424 GTCTACACCGAGGATGACGTGATCCAAACAGCATTTGCGAAGACA----- 32468
 Qy 266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285

Db 32469 -----GGCAACAACACTGGAATTTAATGCTAATGGCACATTA-----GTGATGTGCG 32516
 Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsn 305
 Db 32517 ATGGCGAATATATCGCAACCGCGCAATTAAACGGTGCAGAACCCGCC----- 32564
 Qy 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAsp 325
 Db 32565 -----ACGTTTAGTCTGAGCTTCTCAACTCC----- 32591
 Qy 326 PheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIle 345
 Db 32592 -----ATGCAGCAAAAT----- 32603
 Qy 346 GlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSer 365
 Db 32604 -----ACCGGCGCTAACCAATATT 32621
 Qy 366 ThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGly 385
 Db 32622 GTGGCAAC-----ACCCAGAACGGCTACAAACCCGGGC 32654
 Qy 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 Db 32655 GATCTGGTGTGATTATCAANTCAATGATGACGCTACGTTGTCGCAACTATTCACACGAA 32714
 Qy 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
 Db 32715 CAAACCCCAACGCTGGGGGAGATGTACTGGCGAACTTTGCCAACACGAAAGTCTCGCA 32774
 Qy 426 ArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu 445
 Db 32775 TCCGAAGCGCAACGCTGTGCTGCGACGCAATCTTCTGGCGGCGCTGTGGGACA 32834
 Qy 446 ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAsp 465
 Db 32835 GCCGGGCGGAAACTTTGGCACCTGCACCAACGCTGCGCTGGAAGCGTCCACGTCAT 32894
 Qy 466 MetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLys 485
 Db 32895 CTCAGTAAGAACTGGTCAATATGATCGTCCCGACGTAACATCATCTTAACGCCCGAG 32954
 Qy 486 SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db 32955 ACCATCAAAACCCAGGACGAGATCCTCAACACGCTGTTAACTTACGC 33002

RESULT 31
 AE001509/c
 LOCUS
 DEFINITION Helicobacter pylori, strain J99 section 70 of 132 of the complete genome.
 ACCESSION AE001509
 VERSION AE001509.1 GI:4155350
 KEYWORDS
 SOURCE Helicobacter pylori J99
 ORGANISM Helicobacter pylori J99
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.
 1 (bases 1 to 19549)
 Aln, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., and Trust, T.J.
 Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
 Nature 397 (6715), 176-180 (1999)
 PUBMED 9923682
 REFERENCE 2 (bases 1 to 19549)
 AUTHORS King, B.L., Alm, R.A. and Trust, T.J.
 TITLE Direct Submission

JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA

COMMENT Address all correspondence to: hpearcb.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Jin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astra-boston.com/hpylori). Location/Qualifiers
1. .19549
/organism="Helicobacter pylori J99"
/mol_type="genomic DNA"
/strain="J99"
/db_xref="taxon:85963"
complement(309. .932)
/gene="hsds_2"
/note="synonym: jhp0785"
complement(309. .932)
/gene="hsds_2"
/note="similar to H. pylori 26695 gene HP0849 and HP0849"
/codon_start=1
/transl_table=11
/product="TYPE I RESTRICTION ENZYME (SPECIFICITY SUBUNIT)"
/protein_id="AAD06377.1"
/db_xref="GI:4155365"
translation="MKIERLLTAPKGVGPEKLGEVCESTNKKPTLKI SEYSEVKNK GMYPVNSRELVCYHDFNDNGENTIIASRGYAGFINFYNEKFPAGGLCPYKVKD TNELLTPFLYLNKTNETIMENLVFRGSIPALNKADIEITLPPIPLSIQOEIVTIL DQFSALTDDIQAQGPABEKARKQYREYREKLAFLPPLANKELA"
complement(943. .2529)
/gene="hsdM_2"
/note="synonym: jhp0786"
complement(943. .2529)
/gene="hsdM_2"
/note="similar to H. pylori 26695 gene HP0850"
/codon_start=1
/transl_table=11
/product="TYPE I RESTRICTION ENZYME (MODIFICATION SUBUNIT)"
/protein_id="AAD06378.1"
/db_xref="GI:4155366"
translation="MENTNKASKAKLNHASLANEHLNTIKWVANELRSGVDGWFKQ YVLGIFYRISENAMNHNEYERKIDPNFDYASLSDEAEIVRKSTIEKEGFPPS ALGCVNLKANENDLVANTQNFETIKESLTGPGSEENVKGLFADLDVNSKLGSSH QNRVEKTLTIQAENGWLGDQCSDIGDVFEGDAVEYLMAVNASACKSGEYFTPOEV SELAKITLHOENKYNKYVPCQSGSLLOFSKVLGDKVLKVGFGQISINTLYNLC RINPLHDINSKFHIALGDTILDPRHEDEPDFAIVSNPPYSTKHIGNNNLWNDE RFNKAGALPAADLAFTMUSLYSNSGAAAVBPFGVLIRSGAERKIRYLKEN FIDCVIALPENLFEGTNATCIVLKKNKDDTTLFDISKFEKKGKKNLKAHNRE KILQTYTERKAIHFSAIANIEKIQENDYNLSVRYVEQEDTKIIDIKALNGETSQI VKKSOLRSLSIESI KELEI"
2708. .3394
/gene="jhp0787"
2708. .3394
/gene="jhp0787"
/note="similar to H. pylori 26695 gene HP0851"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06379.1"
/db_xref="GI:4155367"
translation="MVFDRTSVREKKAAXGALGIVGVFFILGVISVASQKWVQ OLDFFDI LRNPAPIQGCTWSLVSFFTWFPAQSCLTTLIALLIGLWFGFQRXALGV WFFTSIILGETLKLALLVAHVDRVNGELVFARGHSPFPSGHALASLYFGSLALL CYANNANIKTGICAILILEWTFILMAYDRVVLGVPSDVLGGFLIGIAWSCCSIALYL

FEATURES source

gene complement(3437. .4376)
/gene="jhp0788"
complement(3437. .4376)
/genes="jhp0788"
/notes="similar to H. pylori 26695 gene HP0852"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06380.1"
/db_xref="GI:4155368"
translation="MKPDIEIIQSVLEAIRKEPIKVTEIDYAKELFKPKGIESMFMDY CGETHPHQSSAAIYTALNKGSELFPKVOEKFPVLIALGAARKEPVINTEKEGVSVKFI VENKIHMSDLHPFLTMYAVYVENLKCVTKITFHBESSKSPKGMORWLYPDWGVSVRIL HAELSNNELAFSKFDTLPKLVSFELKKEISVNCRECYFOALNSSWANEGVLVG RHDTNPNQLMLLXLRLHSFGIGVIDLTBDKSAILLNAKHKEKIDYTVTALESALK NEKPSGLSVVDYDPNHPHYRDEFDEIKKKEELYP"
complement(4375. .5976)
/gene="jhp0789"
complement(4375. .5976)
/genes="jhp0789"
/note="similar to H. pylori 26695 gene HP0853"
/codon_start=1
/transl_table=11
/product="putative abc transporter, ATP-binding protein"
/protein_id="AAD06381.1"
/db_xref="GI:4155369"
translation="MLQTINLTQRYATKKLPENNVIKLDNKNRYGLIGANGAGKSTFL KILKSIDSCSGEVILTSCRMVGLGQDYAFEDLSLXDVLIGNRRLYDALKEKERL YTEGDSLDKVNARIGELETICVEEDPWYCEVIEKLEDLGI PSSXHNLMKTLPS SDFKILLQAVLPFPKPDIILDEPTNNLDLNAIEWENNNKKHGHTVMVISHRHPLN AVCTHLDLPHSVREFSGNYDDYIASTLAKOQEARNKKLEKEKEXFIARFSA NASKAQARSROQLDKLDIOSLAISSRRDSPSII FKPRKTIGNEALECENTISKSYYDO IVLNOYSKVPMPKDKIALIGPNGVGKSTCKILVVELKPDGTGVKNALGVILKGYPPQN VSEIISGBETLQWLFENFNKKIESAEVRNALGRMLFNGEOQSKCNVALSGGEEKHRMVL SKMLSGGNFLVLDPTNHLDELIALGEALFKFDGALICVSHDBRELIDAYANRIIE LVPSPKGASIIDFKGSTEEYFLASKK"
6178. .7155
/gene="guaC"
/note="synonym: jhp0790"
6178. .7155
/genes="guaC"
/note="similar to H. pylori 26695 gene HP0854"
/codon_start=1
/transl_table=11
/product="GMP REDUCTASE"
/protein_id="AAD06382.1"
/db_xref="GI:4155370"
translation="MKVFDYEDVQLIPNKCIVNSRSEDCTTVILGKHAFKMPVPMNM OTTINSIAEFILAENCYFIMRFNGSARI PFMKMKERQLISSISVGVKKEECFLVE ELAKQGTDPDITIDAHGNSVTEMIOQRIKTRLPETFIAGNVYTPVAVRELENAG ADATKVGIGQKVTIKTIGTGTGOWQAALRWCAKARPPIIADGGIRTHGDI AKS IRGATWVIGSLFAGHBSGSETKIENGIAYKEYFGSAGEFKQEKKEIKKIWI HKGSLKNTLIEHMDLQSSISVAGGRDLEAIRKVDYVIVKNSIFNGDTI"
complement(7179. .7757)
/gene="gmbA"
/note="synonym: jhp0791"
complement(7179. .7757)
/genes="gmbA"
/note="similar to H. pylori 26695 gene HP0857"
/codon_start=1
/transl_table=11
/product="phosphohexose isomerase"
/protein_id="AAD06387.1"
/db_xref="GI:4155355"
translation="MIDDLIQEFLAHKEALESLESQDALQSQVHLLIETLENQGH ILICGNGGSADAQHFAELTQYLERKGLSVISLSTDTSALTANDYGVEEFVAR QVEALGNKEDIILIGISTSGNSKNLKA VEKADLQMKTLSLAGRDGKMKPLSDIALI VPSSGDPRIQEWHLMIHLCDIERHPARKN"
complement(7750. .9141)
/genes="waaB"
/note="synonym: jhp0792"
complement(7750. .9141)

Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA
Address all correspondence to: hpearcb.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Jin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: <http://www.astra-boston.com/hpylori>).
Location/Qualifiers
1. .19549
/organism="Helicobacter pylori J99"
/mol_type="genomic DNA"
/strain="J99"
/db_xref="taxon:85963"
complement(309. .932)
/gene="hsds 2"
/note="synonym: jhp0785"
complement(309. .932)
/gene="hsds 2"
/note="similar to H. pylori 26695 gene HP0849 and HP0849"
/codon_start=1
/transl_table=11
/product="TYPE I RESTRICTION ENZYME (SPECIFICITY SUBUNIT)"
/protein_id="AAD06377.1"
/db_xref="GI:4155365"
/translation="MKIERLLTAPKGVGPEKLGVCESNTKKPTLKITSEVSEVKNK GMYPVNSRELGYHDFNDGENTIIASRGYAGFINFYNEKFPAGGLCPYKVKD TNELLTFLYLNKTNETIMENLVFRGSIPALNKADIIETLTIPIPLSIQOEIVTIL DQFSALTDDQAGIPAEIKARKQYREYREKLLAFKPLPLANKELA"
complement(943. .2529)
/gene="hsds 2"
/note="synonym: jhp0786"
complement(943. .2529)
/gene="hsds 2"
/note="similar to H. pylori 26695 gene HP0850"
/codon_start=1
/transl_table=11
/product="TYPE I RESTRICTION ENZYME (MODIFICATION SUBUNIT)"
/protein_id="AAD06378.1"
/db_xref="GI:4155366"
/translation="MENTNKASAKLNNHASLARNELHTIWKVANELRSGVGDWFKQ YVGLIFYRISENAMNHNEYRKDPNDYASLSDEAEIVRKSTIEKEGFFPPS ALKCNVLNKNENDLVNTQNTFNETKRSGLTPEENKVGFLADLVNSKLGSSH QNRKELTKILOAGWQMGDQCGSDIGDVFGDAYEYLMAYASNAGKSGEYFTPEV SELAKITLHOENKYNKVPQCGSGLLOFSKVLGDKVLKGVFGQELINTVNLK RINPLHDINSKHIALGDTLDPHREDEPDALVSPPYSTKHWGNPLNNDDE RFNKAGALNADADLAFTMUSLYSNGAAIVBPPGVLYRSGAEKIRLYVKNR FIDCVIALPENLFGTNIATCIVLKNKQDDTLTDFIDASKFLKGGKKNKILAHNRE KILQTYTERKAIKHFSALEIKETQENDYNSVRYVEQEDTKIIDIKALNGEISQI VKKQSVLRNSDESIIKELEI"
2708. .3394
/gene="jhp0787"
2708. .3394
/gene="jhp0787"
/note="similar to H. pylori 26695 gene HP0851"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06379.1"
/db_xref="GI:4155367"
/translation="MVFDRTSVREKKAAGALGIVGVFFILGVISVASQKQWVQ QLDFFDILRNPAIQGTSWLSVFTWFAQSKLTTLIALLIGLWFGFQRIALGV WFFSIIILGFTLLKLLVLRPNVNGELVFHAGFSPPSGHALASLYFGSLALL CVYANNNAIKTIGAILILEWTFILMAYDVRVILGVYPSDVLGGFLIGLAWSCCSIALYL

```
/gene="waaB"  
/note="similar to H. pylori 26695 gene HP0858"  
/codon_start=1  
/transl_table=1  
/product="putative ADP-D-glycero-D-mannoheptose synthase"  
/protein_id="AAD06368.1"  
/db_xref="GI:4155356"  
/translation="MKKILVIGDILADYLMWKSERLSPAPVPVLVLEKESKQLGGA  
ANVANNITSLKAKVFLGVGDDLEKGFSTLTKRGIDSGVLIDKTKTLKTRII  
AQNQIVRVKKEIKDPLNADLRNKLDPFAEKIQEIDGVILSYNKGVLDFELTQRII  
TLANKHKLILCDPKGDIYKSHASLITENRAELQALHKLQDSHANLSKALQIQLE  
TYHAMPVLAASVVGKMGALASLEBIALILNTHPKILPLEKLETLERNQOKTVF  
KEACFALHKGHASYLQAKALGDIILVGLNSDINSIKRLKGRKPITVSKDRAPFLAS  
TNGCFDILHKGHASYLQAKALGDIILVGLNSDINSIKRLKGRKPITVSKDRAPFLAS  
LSCVDYVVVFGEDETPIKLIQALKPDIILVGLNKEVIGSELAKETRLIEFEGYST  
SALIEKIKRTHND"  
complement(9138..10127)  
/gene="gmdH"  
/note="synonym: jhp0793"  
complement(9138..10127)  
/gene="gmdH"  
/note="similar to H. pylori 26695 gene HP0859"  
/codon_start=1  
/transl_table=1  
/product="ADP-L-GLYCERO-D-MANNOHEPTOSE-6-EPIMERASE"  
/protein_id="AAD06369.1"  
/db_xref="GI:4155357"  
/translation="WRVYIDGLENQITLITGGAGFVGSNLAIFYQENHPKAKVVLDDK  
FRSNTLSNNRPSLSLGHFKLIGFKGEVIADINPLDLRLLEKLFDYLFHQAAVSD  
TTMLNQELVMKTYQAFNLLEIARSKAKVIYASSAGVYGNTKAPNIVGSNESPENV
```

Alignment Scores:
Pred. No.: 3,34e-21 Length: 19549
Score: 485.00 Matches: 181
Percent Similarity: 38.04% Conservativity: 94
Best Local Similarity: 25.03% Mismatches: 222
Query Match: 18.75% Indels: 226
DB: 1 Gaps: 22

US-10-009-823A-1 (1-502) x AE001509 (1-19549)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
Db	19415	ATGCTAGGCTTTATGCTGTGTGTCATGATGGATGCAAGCCACCAACCGCTTTGGAT	19356
QY	21	ThrValSerAnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnVal	40
Db	19355	ATTGAGAGTACAAATATGCGAAGTAAATACCACTGGGTTTAAAGTATTCAGGGCTTCT	19296
QY	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer	56
Db	19295	TTTGTGATAGCTCTCTCAAGTCAAACTCATGCTACCCACCCATATAAAACGGGTG	19236
QY	57	GlnGlyProAn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle	74
Db	19235	CGCGGGCAGAGTACTTTCTTGTGGGCTTGGGTAGCGGTGGATGCGACGACTAAATC	19176
QY	75	PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys	94
Db	19175	TTTTCGCAAGCAATATCCAAACACCGATGTCMAAACCGATTTAGGATTCAGGCGAT	19116
QY	95	GlyPheGlnValThrLeuGluAspLysVal-----HisTyrThrArgAlaGlyAsn	112
Db	19115	GGCTTTTATCATTCAGCTGATAGGGGATCAGCGCAATTCACGAGATGGGAG	19056
QY	113	PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet	130
Db	19055	TTTCTTTTGTACTGCAAGGAGTTTGTGTACACCGGTGGCTTGTGGTGCAGGGTGG	18996
QY	131	-----GlySerArgIleSerAsnAsnProAsnIleLysLysGlu	143
Db	18995	GTGAGAAACGGGACGATACCGGTAAATAAGGAGCGATACGCGACGCTTTAAAGTGGAT	18936

QY	144	ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLys	163
Db	18935	AATACCGGCCCTTTAGAAAACATAGGATTCAGCAAGATGGTG--ATGCCTGTGTA	18879
QY <th>164</th> <th>ThrSerThr-----AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr</th> <th>181</th>	164	ThrSerThr-----AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr	181
Db	18878	GGCAGTAAACCCGATTTCTATGAGGCGCAATTTAAACGCTGGAAGCATGCGGATCAACA	18819
QY <th>182</th> <th>GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLys-----GlyAsnGly</th> <th>199</th>	182	GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLys-----GlyAsnGly	199
Db	18818	GGCGGG-----GTATTCGCTTAGATTCCTTCAGCCAAACCCCTTCAGATGGC	18771
QY <th>200</th> <th>ThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArg-----</th> <th>215</th>	200	ThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArg-----	215
Db	18770	ATTAATCCAGTGTATGATTCAGGCACGAATCTAGCTCAAGTCGCCGAAGCATGGGATCT	18711
QY <th>216</th> <th>ValTyrAspGlnGlnGlyAsn-----SerHisAspIleThrValTyr</th> <th>229</th>	216	ValTyrAspGlnGlnGlyAsn-----SerHisAspIleThrValTyr	229
Db	18710	TTATCAATGAGATGGCGACGCTCTTTTATTGATGAACCAAGGATTTGGTGAGC	18651
QY <th>230</th> <th>PheAspGlyAla-----ProSer-----</th> <th>235</th>	230	PheAspGlyAla-----ProSer-----	235
Db	18650	TATAAGAGTGGCAAAATGGTTAAAGACATCTCTCTTCGAGAAACAGCACGCTTGAG	18591
QY <th>236</th> <th>-----SerThrGlySerLysThrPheGluTyr</th> <th>244</th>	236	-----SerThrGlySerLysThrPheGluTyr	244
Db	18590	TTGAATGGGTAAAGATTTCTTTCAGAACGATTCAGCGGTGAGCAGACTTCAAGCTTA	18531
QY <th>245</th> <th>LeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly-----Thr</th> <th>260</th>	245	LeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly-----Thr	260
Db	18530	GTGGCGGCTAAATTCGATCAATCGGTTAAAGCCAAACAGCAGCTTGAAGCTTATTTG	18471
QY <th>261</th> <th>AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys</th> <th>280</th>	261	AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys	280
Db	18470	GACGCAAGCAATTTGCGTTTGGAAACACCAACGAATTAGACGCGGATGAAGACCTTAA	18411
QY <th>281</th> <th>Asn-----</th> <th>281</th>	281	Asn-----	281
Db	18410	AACATTGGTCTCACTCAAGCCGGACCGGCGTTCGCTAACTTTTATAGCGGCGATAAA	18351
QY <th>282</th> <th>---MetThrAlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsn</th> <th>296</th>	282	---MetThrAlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsn	296
Db	18350	GATGTAACGGCTTTCAATACATAGTATATACGATCTATTATGCGGATGATTTGGG	18291
QY <th>297</th> <th>AlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheVal---</th> <th>315</th>	297	AlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheVal---	315
Db	18290	CAGTTTAGACCACTGAAGACTTGGCGCTTATCCAGCATGACGCTATATCGTTAAA	18231
QY <th>316</th> <th>-----GlyAlaGlyIleGlnProLeuThrLeu</th> <th>324</th>	316	-----GlyAlaGlyIleGlnProLeuThrLeu	324
Db	18230	GATCCTAGTTTAGCGACAATTTCAAGATTCAGCGCTTCTATAGGGGTTCACATCAAC	18171
QY <th>325</th> <th>AspPheGly-----IleLysSerGlnGln</th> <th>332</th>	325	AspPheGly-----IleLysSerGlnGln	332
Db	18170	CAATACGCGATGTTTGAATCAACATAAGATTAATAAATGTCNTTAAGAAAATCTC	18111
QY <th>333</th> <th>AsnMetTrp-----</th> <th>335</th>	333	AsnMetTrp-----	335
Db	18110	AATATCTTTGTAGTGGGTATTTCTTCAGACAGCGTAACGAACAATGTTTGTATAAAAC	18051
QY <th>336</th> <th>-----AlaGlyAlaProAlaSerAlaAla</th> <th>343</th>	336	-----AlaGlyAlaProAlaSerAlaAla	343
Db	18050	GCCATGAAGTCTTAATACCGCTTCTTTAATGAAGGGGAGCGTCAGCTAGCAGTTCT	17991
QY <th>344</th> <th>-----AlaIleGlyThr---</th> <th>347</th>	344	-----AlaIleGlyThr---	347
Db	17990	AAATTCACCAAGCTACCCATCGCAAGCATTGATGTATAGACAGTTTATAGGCACTAAA	17931
QY <th>347</th> <th>-----</th> <th>347</th>	347	-----	347
Db	17930	CACGCCATGCGCATTGAGTTTTTATAGGAGTGGGGGGGGAATGGAATTTTAGAGTGCATC	17871
QY <th>347</th> <th>-----</th> <th>347</th>	347	-----	347

```
Db 17870 GTGCTGAGCTGGGGAATTAGTAGGGGGTCTAGCGGCTAGGCTAATGTGTGTAAGGG 17811
Qy 348 -----AspIleGlyLysLeuProSerMetMetPro----- 357
Db 17810 GCGCGTTGCAATTTAATAACGCGGATCGCTTCGGGCGATGAACCGGCTCTTTTGCAA 17751
Qy 358 -----IleGlnThrSerSerGlyAsnSer 365
Db 17750 TTTGACCTAAAATGCGCGTGATGCCCCCAACGCGATCAATTAGCTTTTGTTCTCTCA 17691
Qy 366 ThrAlaArgAsnGlySerSerThrArgArgTyrSer-----Gln 379
Db 17690 GGGAGCTTTGACGGAATGACAGCGGTGATAGATTTCTGAAACTTATCGATTGAGCAA 17631
Qy 380 AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
Db 17630 AACGGCTATCAAGCGGCGGATTTGATGATGTCGCTTTGATTTCTGATGGGTGCTTTTA 17571
Qy 400 GlyLysTyrSerAsnSerGlnValAspPheTyrAsnIleProLeuAlaArgPheThr 419
Db 17570 GGGCGTTTCAATGATGCGAGGACTTTAGCGCTCGCTCAAGTGGCTTAGCGAATTCGCT 17511
Qy 420 SerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db 17510 AACGATGCGGCTTACAGGCTTTGGAGGCAATGTCTTTCTCAAAACCGAAACTCTGGG 17451
Qy 440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db 17450 CAAGCTTAATCGTGGCGTAATACGGCGTAGGGGTTCATTTTCAGATCCAAACTA 17391
Qy 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGly 479
Db 17390 GAGCTAGTAATGCGATTGACGCGGAGTTTAAACGAAATTGATTGCTGTTCAAGGGGG 17331
Qy 480 PheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
Db 17330 TTTCAAGCGAACTCTAAAGCGGTAAACCATCCGATCCAAATCCTTAACCCCTATTGAAT 17271
Qy 500 LeuLysArg 502
Db 17270 CTCAGCAA 17262

RESULT 32
AC095028
LOCUS
DEFINITION
Leishmania major chromosome Unknown clone P864 strain Friedlin, **
SEQUENCING IN PROGRESS ***, 4 ordered pieces.
AC095028
AC095028.2 GI:117352423
HTG; HTGS_PHASE2.
Leishmania major
Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 90250)
Myler,P.J., Sisk,E., Cawthra,J., Nelson,S., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H., Stuart,K., Mack,J.,
Marty,A., Rinta,J. and Seyler,A.
Direct Submission
Submitted (15-SEP-2001) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
On Dec 5, 2001 this sequence version replaced gi:15624885.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 7210: contig of 7210 bp in length
```

FEATURES

source

```
* 7211 7411: gap of unknown length
* 7412 27989: contig of 20578 bp in length
* 28189: gap of unknown length
* 28190 58975: contig of 30786 bp in length
* 58976 59175: gap of unknown length
* 59176 90250: contig of 31075 bp in length.
```

Location/Qualifiers

```
1..90250
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="Unknown"
/clone="p864"
/note="This PAC appears to contain bacterial DNA. It is
not yet clear whether it is derived from a contaminant or
a symbiont."
```

ORIGIN

Alignment Scores:

```
Pred. No.: 2 05e-20 Length: 90250
Score: 485.00 Matches: 153
Percent Similarity: 39.82% Conservatives: 74
Best Local Similarity: 26.84% Mismatches: 177
Query Match: 18.75% Indels: 166
DB: 2 Gaps: 19
```

US-10-009-823A-1 (1-502) x AC095028 (1-90250)

```
Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla 27
Db 80344 GGAGTCAGCGGCATCAATGCCGCGCACAGAAATCGGACGTGATCGGCAACACATCGCC 80403
Qy 28 AsnAlaSerThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln 47
Db 80404 AACGGAAACCGTCGGCTTCAAGCAATCGACCGCGAGTTCCGCCGAGCTCTACGCC--- 80460
Qy 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
Db 80461 -----GGCTCG-----AAGGTCGGCTCGGCACGCGC 80487
Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
Db 80488 GTGAACTCGGTGGTGCGAGTCGTTTCCAGCAGGCGCAATATCGAGCGAGCGCGCTCGCTG 80547
Qy 88 AspLeuAlaIle---GlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysVal 105
Db 80548 GACGTCGCGATCACCGGGGCAACGGCTTCTTCGCGATGTTCCGCCACCGCGCTCGGTG 80607
Qy 106 HisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer 125
Db 80608 TTCTTCTGGCGCAACGGCGCAATTCCTGCGCACCGACGCGCGCATCACCAACGCGCAG 80667
Qy 126 GlyPheThrLeuMetGly----- 131
Db 80668 GGCTGAGCTGACCGGCTATCCGGCGGGGTGCGAGCGGGCGCGGCTGCGACCGCAA 80727
Qy 132 -----SerArgIle----- 134
Db 80728 CCGCTGGTGGTGTCACGCGCCAGATGCGCGCGAAGGCCACCGACCGCATCACCGCGCAA 80787
Qy 135 -----SerAsnAsnProAsnIleLysLysGlu----- 143
Db 80788 TACAACCTGGACTCGCGCGCGCTTCTCGCAATCCAGCTGCGCGCAACCGCGCTAAT 80847
Qy 144 -----ThrLeuGluProIle--- 148
Db 80848 ACGGGATCAGGTTTCGGTACAGGTTCGCGAGTTTCGACGCAACCGCGCTGAGGACTTCACC 80907
Qy 149 ---GlnLeuAspPheAsn-----AspProThrValAlaLysSerProAlaLysThrSer 165
Db 80908 TACCAGCTGTCTTCGATGGCACCGATATACACCTGACCCCTTCTGTATGGGCAACGCTG 80967
```

```
Qy 166 ThrAlaLeuAsnAlaVal-----Val 172
Db 80968 ACTCAGCGGGCAGCAGCTTACGATAGCGGCTGACGCTGACAGCGGCGCGGTATG 81027
Qy 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192
Db 81028 AACACCGGTGACCC-----TTTACCGTCGATCGCTCACCCTCACTTC 81072
Qy 193 GluSerTrpLysGlyAsnGlyThrProPheLysThrSerAsnTyrSerTyrAlaGln 212
Db 81073 ACTAGTCGCGCGGCAAT-----CCGCTCGCAGCAGAACTGCTCACTACACGAC 81123
Qy 213 ProMetArgValTyrAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232
Db 81124 ACGATCAGGTGTCGACTCGCTCGGCAATGCCAGCAGTCAAGCCTATTTCGTCAG 81183
Qy 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252
Db 81184 GGTCCGACCGCGGGAACGTCAGCGCTGGATGTCTACGTG----- 81225
Qy 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly-----Thr 270
Db 81226 -----ACGACAGCAGCGCGCGCCCACTGCTCGGTCCCAATGAG 81264
Qy 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
Db 81265 ATGGTGTTTAACTGACGCGCGCCTG-----GTAACCTCGCGCGCGCGC 81309
Qy 291 AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe 310
Db 81310 GCAACG----- 81315
Qy 311 SerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSer 330
Db 81316 -----ATCCCTGTGACCGTC----- 81330
Qy 331 GlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGly 350
Db 81331 -----CCCGAGTCCGCGCGCGCGCGCGC----- 81354
Qy 351 LysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGly 370
Db 81355 -----ATGAACATCCAGCTGAGCCTGCGCGCGCGCGCGCGCGC 81399
Qy 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
Db 81400 GACACGATCCGGAAGTCGATCGTCGACAGCGGCTATACCTCGGCTCGCTCGGCGCTC 81459
Qy 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
Db 81460 TCGATCGAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81519
Qy 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyAsnAsn 430
Db 81520 GGTGAGTGGTGAGGCTCGTTCGCAACGCTCGAAGGCTCAAGCCGAGCGCGCAAC 81579
Qy 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db 81580 GTCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81636
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 81637 ATGGCCAGCTGAGTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81696
Qy 471 ValAsnMetIleIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
Db 81697 GTCAACCTGATCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81756
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeu 500
Db 81757 GACAGGGTCTGCAGACGCTGTCACATC 81786
```

```
AL591976
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
```

```
AL591976 250050 bp DNA linear BCT 06-JUN-2002
Listeria monocytogenes strain 53D, complete genome, segment 4/12.
AL591976 AL591824
AL591976.1 GI:16409987
Listeria monocytogenes
Listeria monocytogenes
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloecher, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussutget, O., Etian, K. D., Fsihi, H., Portillo, F. G., Garrido, P.,
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L. M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F.,
Kurapat, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E.,
Nedjari, H., Nordstiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C.,
Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N.,
Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
21337279
11679669
2 (bases 1 to 250050)
Glaser, P., Frangeul, L. and Rusniok, C.
Direct Submission
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
E-mail: pglaser@pasteur.fr
Phone: +33 1 45 68 89 96, Fax: +33 (0) 1 45 68 87 86.
Location/Qualifiers
1. 250050
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/strain="53D-e"
/db_xref="taxon:1639"
/complement (116..121)
/gene="lmo0598"
/complement (116..121)
/gene="lmo0598"
/join (287..292,300..632)
/gene="lmo0599"
/gene="lmo0599"
/gene="lmo0599"
/gene="lmo0599"
/notes="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="CAC8677.1"
/db_xref="GI:16409988"
/db_xref="TrEMBL:Q92E53"
/translation="MEVNPQFKGVLELCLFLQKDCYGYELANQVSKYIEVAGEA
IYVLRRLVKEEYCYSTLYVESNPGSRKYLTVKGBIYLNELISEWNNFTDSVAKLL
TEGEAVNE"
610..615
/gene="lmo0600"
610..615
/gene="lmo0600"
625..1215
/gene="lmo0600"
625..1215
/gene="lmo0600"
625..1215
/transl_table=11
/protein_id="CAC8678.1"
/db_xref="GI:16409989"
/db_xref="TrEMBL:Q8Y9D2"
/translation="MNQDFLNLNQRLLELDPKERRELIASDYQEHFNGIEAGSBE
QIVFDLGPKEIAADIISERGIREPTMDYVYVPRKNQENRSVSKILIGVGLFLL
```

```
DVCLIPIMVSLVSLVLSLWTVGAFLLSPVLGVGLFGADFEYQMFVSGVLGLG
LMLLFVANALQTLTSKVTWAILQWHKVIYKGGKNA"
/genes="lmc0601"
1193..1198
RBS
/genes="lmc0601"
1193..1198
gene
/genes="lmc0601"
1208..2308
CDS
/genes="lmc0601"
1208..2308
/genes="lmc0601"
/notes="similar to cell surface protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98679.1"
/db_xref="GI:16409990"
/db_xref="EMBL:Q8Y9D1"
/translation="MKHKKSLKFLFAGLVLFIIIGAVFMTMTGKMEKEPELTKQ
WLSITENIKKIFSSERDASIEWKESTTKYIELKGNYSANDKKAIOQLPEVSDGT
SPDITVPEDDWNGFKIYAYKQKQVTLTKTKLADLEVYKSHSGIDIVADFKVK
FVSTNSGKLVNLEANTAOATSGDITLGNKANGSIETDSKGTETLNTLGTLEV
NGSGDVNAGVAKKLTAKIDSGDIELTSGVTVDLAVITTSGGDIDANTKGIQARS
DGSIELACATNNVAKTSSGDIIDVAFKQVKNIEINSDSGEVELELPQDFKAIYAS
SNQSGVAKPTSDSNDTNRVTVKTSDDIKIEK"
join(2400..2405,2414..2914)
/genes="lmc0602"
2400..2405
RBS
/genes="lmc0602"
2414..2914
CDS
/genes="lmc0602"
/notes="weakly similar to transcription regulator"
/codon_start=1
/transl_table=11
/protein_id="CAC98680.1"
/db_xref="GI:16409991"
/db_xref="GOA:Q8Y9D0"
/db_xref="EMBL:Q8Y9D0"
/translation="MDFHKKATNSDAEAIQHVATSHHTYQDLIPGDVQDDFLERF
XNVETLHNRISATPFVAVQADKVGIFANFIELEKKSLEAAFPYLLPEVTQKGLTEL
LEVGMTLFVPLPMFVNVKGNETAHFYKAKGFQVVEFTDFYGYLETIRFLNHN
HAPEEP"
gene
complement(2909..2956)
/genes="lmc0603"
terminator
complement(2909..2956)
/genes="lmc0603"
2920..2967
gene
/genes="lmc0602"
2920..2967
terminator
/genes="lmc0602"
complement(join(2962..3348,3356..3361))
/genes="lmc0603"
complement(2962..3348)
/genes="lmc0603"
/codon_start=1
/transl_table=11
/protein_id="CAC98681.1"
/db_xref="GI:16409992"
/db_xref="EMBL:Q8Y9C9"
/db_xref="EMBL:Q8Y9C9"
/translation="MRPIIAYUSIFVLGIFSAALLVETIYDNNVTQVLSAILFAAPV
ILVATIGIFYGFSGKASYFTFAIMGFAYGVAIVLIIQVSGMLISVGSVILAG
VIMALLAIIPFFLRGGKSTSGKAATK"
complement(3356..3361)
RBS
/genes="lmc0603"
complement(3400..3761)
gene
/genes="lmc0604"
complement(3400..3744)
CDS
/genes="lmc0604"
/notes="similar to B. subtilis YvIA protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98682.1"
/db_xref="GI:16409993"
/db_xref="EMBL:Q8Y9C8"
/translation="MLDKKIISALYSFSLFFAFVITPAIINWCTTKDKIRHHVKWALL
```

```
TOATFVAGIVVMILLNNVPFSTNSADTINFLATVFFVFLNVSILFNLIRGITR
IVKSDDNARC"
complement(3756..3761)
RBS
/genes="lmc0604"
complement(3891..5244)
gene
/genes="lmc0605"
complement(3891..5231)
CDS
/genes="lmc0605"
/genes="lmc0605"
/notes="conserved hypothetical membrane protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98683.1"
/db_xref="GI:16409994"
/db_xref="GOA:Q8Y9C7"
/db_xref="EMBL:Q8Y9C7"
/translation="MSSVHKVASLSLFTLAWPIFLFLEQLMISYIDVFMGLHYSDDA
VATGVAQNQILVISIIYGFISVGVQIIIVAOIGAKKHEIENVITNGVAVLIGIV
MSIIFPMSKNFLTMGIDPHLVQVGAPFIEIIGGSSVTAIHASILPILRAKGYVRQ
SILIPVTISINNVGNLYFLYPLAYLDYGVAGVISTAVNFVGMGLAIWMLRKYIG
YTFHFKLEQVSKKLYSILRLGSPAGENSAGVQVVTALAILGTEALTKVYA
STVQFVALFAIALGQASQIIIGRAVGAKEDIKAYKGLSWKIGLVVAIVVSYSYL
FABPIMLFTTNTETIANTKELFLLSIFLELRATNIIISLSNSTGDVFPFTCGLI
VMNIVSPFSYVLGISAGLGVGLWAYIIDEGVRAVLMYRRWSKWSLKSVI"
complement(5239..5244)
RBS
/genes="lmc0605"
5384..5858
gene
/genes="lmc0606"
5384..5389
RBS
/genes="lmc0606"
5397..5858
CDS
/genes="lmc0606"
/notes="similar to transcription regulator MarR family"
/codon_start=1
/transl_table=11
/protein_id="CAC98684.1"
/db_xref="GI:16409995"
/db_xref="GOA:Q8Y9C6"
/db_xref="EMBL:Q8Y9C6"
/translation="MSSKNQDLGHSVTKAFNFKHAIKSFQIPGYKSETRFIFILS
RGLSKRGPRIKRVSDLGHMLRISKPSVTOMIOSLEGKGLIKRVQNPDKRSMYVELTEV
GAGVSEKMLDFEQASFEDMQEFLGEDDMKKLITLLEKLTDLNTKSENKEA"
join(5865..5871,5878..7590)
gene
/genes="lmc0607"
5865..5871
RBS
/genes="lmc0607"
5878..7590
CDS
/genes="lmc0607"
/notes="similar to ABC transporter, ATP-binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98685.1"
/db_xref="GI:16409996"
/db_xref="GOA:Q8Y9C5"
/db_xref="EMBL:Q8Y9C5"
/translation="MKLKPVWLSITAVLVLTGQVIGQLYLPFLMSNIIDKGVVTGD
```

Alignment Scores:

Pred. No.:	7,38e-20	Length:	250050
Score:	484.50	Matches:	137
Percent Similarity:	42.21%	Conservative:	77
Best Local Similarity:	27.02%	Mismatches:	188
Query Match:	18.74%	Indels:	105
DB:	1	Gaps:	11

US-10-009-823a-1 (1-502) x AL5191976 (1-250050)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetIysThrHisSerThrGlyLeuGly	20
Db	90516	ATGAATCAAACTATGATACAGCTATTTCGGATGAATCGTTCACACAGCATATCA	90575
QY	21	ThrValSerAsnAsnIleAlaAsnAlasThrIleGlyTyrLysGlnGlnValVal	40
Db	90576	GTAACATCAATATATTTCCTGCAATGCCACACAGCATATATAAACAAGCTCGT	90635

QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
 DB 90636 TTCAATGATTACTTTTACCAAAACACAAATGGGATCTGTTCCAGCGGACATTTATGCTGGA 90695
 QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 DB 90696 ACAAAACCAATGATTCGTTCCGGTTCGAAATTCGGCGATTTAACTGACTATACA 90755
 QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlySerGlyPhe 96
 DB 90756 GCAGGTTCCCGACATTCACATGCGCAGAAATAAAGATCGACCACTACAAGCCCGCGCTTT 90815
 QY 97 PheGlnVal-----ThrLeuGluAspLysValHisThrThrArgAlaGlyAsnPheArg 114
 DB 90816 TTCAATGCTGTAATACGCTGGTGGGAATATCGTTTATACGTGACGTGACTTTTGA 90875
 QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
 DB 90876 GTATCTGACAATACTATTAAACACGACGCAAGAAATACGTTATGCGTTACGCAACG 90935
 QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 DB 90936 GACAAAATGTAACGTTTAAACGGTAATTTACACCCATCCAAATTCGCTAATATAGC 90995
 QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
 DB 90996 GCAATTCAGGAGCAACACAAAATGTTAGCTAAGCGGTAACTCCACATGATGTTGG 91055
 QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
 DB 91056 GCGCAAAAA---GATACGATTTCTTCGAG----- 91082
 QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
 DB 91083 -----CTG 91085
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
 DB 91086 TCTGTATACGATAATGCTGCGGAAAAACATAAACTTCAAGTCAATATGAAGCAGCTACA 91145
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 DB 91146 CCAGATCGAGCGTAATGTTCTAGCAATATGAATTCAAATG-----GAC 91193
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 91194 GGCAGGACATTAATCTCTCCA-----GTGACAGGACACTTAATTATAC 91235
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 91236 AATGCGAGGCTGAATCAACGACCAACCCAGACGCA----- 91268
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 DB 91269 ---CTTAAATAATTCAAATCAATTCACAGTAAACGGC----- 91304
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 91305 -----AAACAGTCAAT 91316
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 DB 91317 ATG-----GGCTTAACCTAAGTGGCTTAACA 91343
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
 DB 91344 AACTAC-----GGAACGACCAAA 91361
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 DB 91362 GTATCTCACCAACTTCTGTATGCAAGGCTGCGACTGTAAAGATTTATGCAAGTTATAC 91421

QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 DB 91422 GATTCGTGCTATATTGCAGTGAAGTATTTCAGATGGTACAGTATCCCGAGTGGCCCACTC 91481
 QY 414 ProLeuAlaIaa-gPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
 DB 91482 GCGGTTGCTACGTTCTCCCAATGAAGACGGTTTACTCAAAATGGGGAACGGCGAATATGTT 91541
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 DB 91542 CCAGGATTTATCTTCTGGFGATGCGATGACGGCTGTCGCCAAATGGCGCTGGCGGA 91601
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 DB 91602 ATTAGTGTTCTTCTCCAGAGGATCAACGATAGACTTGTCCTGGAATTCGTTAACTTA 91661
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 DB 91662 ATGACATATCAAGTGGTTTCCAAAGGAATACAAAGGTTATTCGTGTCGGGATGACGTG 91721
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 DB 91722 ATGAACAAATTTGAACTTG 91742

RESULT 34
 LOCUS AL596166 260050 bp DNA linear BCT 06-JUN-2002
 DEFINITION Listeria innocua Clip11262 complete genome, segment 4/12.
 ACCESSION AL596166 AL592022
 VERSION AL596166.1 GI:16413128
 KEYWORDS Listeria innocua
 SOURCE Listeria innocua
 ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

REFERENCE
 1 Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P., Donann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dusurget, O., Entian, K.D., Fsihi, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kref, J., Kuhn, M., Kunst, F., Kurapat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordisk, G., Novella, S., de Pablo, B., Perez-Diaz, J.C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tisserat, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P. Comparative genomics of *Listeria species* Science 294 (5543), 849-852 (2001)
 21537279
 MEDLINE
 PUBMED 11679669
 2 (bases 1 to 260050)
 Glaser, P., Frangeul, L. and Rusniok, C.
 Direct Submission
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 E-mail: pglaser@pasteur.fr
 Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
 Location/Qualifiers
 1. 260050
 /organism="Listeria innocua"
 /mol_type="genomic DNA"
 /strain="Clip11262"
 /db_xref="taxon:1642"
 331..1225
 /gene="lin0677"
 331..336
 /gene="lin0677"
 344..1225
 /gene="lin0677"
 /note="conserved hypothetical protein"
 /codon_start=1
 /transl_table=11

gene
 RBS
 CDS

FEATURES
 source

REFERENCE 2 (bases 1 to 290242)
AUTHORS Nelson,K.E., Fouts,D.E., Mongodin,E.F., Ravel,J., DeBoy,R.T.,
Pasko,D.A., Kolonay,J.F., Angiuoli,S., Gill,S.R., Paulsen,I.T.,
Peterson,J.D., White,O., Nelson,W.C., Niernman,W.C., Van Aken,S.E.,
Khouri,H.M., Fedorova,N.B., Forberger,H.A., Iran,B. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2004) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 290242
/organism="Mistaria monocytogenes str. 4b F2365"
/mol_type="Genomic DNA"
/strain="4b F2365"
/db_xref="taxon:265669"
83. .1354
/locus_tag="LMOf2365_0566"
83. 1354
/locus_tag="LMOf2365_0567"
/notes="identified by similarity to SP:Q01264; match to
protein family HMM PF01546; match to protein family HMM
TIGR01879"
/codon_start=1
/transl_table=11
/product="N-carbamoyl-L-amino acid amidohydrolase,
putative"
/protein_id="AA03348.1"
/db_xref="GI:46880048"
translation="MYQNIKQAVLNNEEAVIAPRDLNHPPELQWQEFRTTKVAKEL
ENAKVLAVSEDAIGNYGRLEGENPNAIVGVHFDSPVNGGAFDGPAGVITGLV
ATVYEQQIKPHEPLRIIAVBEEGARFGALLARTITGKVTKEMLHEMKDIDGITA
AEAMKGFDPANOVTAITRESKVAFLIHEIQGPVLNANEDVALDVTGVLTEIK
VTVKGAGHAGTTPMLRDALITAVSIIIGQLPELAIOEGGCTVTVGKLVNYPNGA
VIPKVVFTDIPRAKDEIHVQNTLAKTEIIIOAEKNGITCEIEBMLPEPPLHLSKEI
HQALTESDOLGLKRYTSGAGHDAMIAGTVEGLVIFVPSHNGISHAPEWTDYDK
LQKGIKIVLETKVKTTEESANEKNTSSFK"
1320. .2495
/locus_tag="LMOf2365_0567"
1320. 2495
/locus_tag="LMOf2365_0567"
/notes="identified by similarity to SP:P80092; match to
protein family HMM PF01546; match to protein family HMM
TIGR01891"
/codon_start=1
/transl_table=11
/product="carboxypeptidase, putative"
/protein_id="AA03349.1"
/db_xref="GI:46880049"
translation="MNQIKQAVLNNEEAVIAPRDLNHPPELQWQEFRTTKVAKEL
DKLGIPYRTEPTGLIAELKGGSGKTVLRADMDALPVQELNQLDPYKSTEDGKQHA
CGHDAHTAMLTAAKALVEIKDELPGTVRFIPOPSEIIEAGAKAMIAQAGMEGVDRVF
GIHWSTPSGKTSCVVGSTFASADIIQIDPKGGGHPHDITDAAVIASFVNNL
QSIYSRETDPLDPVVVTIGKMDVGYTRYIAENARLEGTLRCFNNITRAKVAKSTEHY
AKQTAALYGGKAEIMYKQGTQPVINDEKSAALLVQETITSPGCEMLYFERPTGGEDF
SYFDAPGSPFALVGCNPEKOTEMAHHGRFNIDESVMKNGHLYAQFAYNPLQN"
complement (2543. .3559)
/locus_tag="LMOf2365_0568"
complement (2543. 3559)
/locus_tag="LMOf2365_0568"
/notes="identified by match to protein family HMM PF04274"
/codon_start=1
/transl_table=11
/product="tagatose 1,6-diphosphate aldolase, putative"
/protein_id="AA03350.1"
/db_xref="GI:46880050"
translation="MVQITKGGDLQRLNKEGVIAALAIQDGLSKKMIQOAKGTE
NKQVEDPKQLVSEELTPYSAILLDLGYCTPAIKARHEGSGLLTSYKTYDATTPG
KPLDILDLNARIKENGGDAVKILVYDDEPAEINEIKYAELEIRGAECDADIPF
FLEPTIDATVDSGLUEYAKLPKARKYKIFSPKRYGVDTVLKLEVPNFKYVEGF
AEGVATYQDEAAHFECDLGPLPIYLSAGVTSEMFKHTIQFANQHNQVSGVIG
GRWADGIEYVGGQGDALREWLRTQKENITSLDKLDELGAVPWMTKYGSPEDVHV
VEKO"
3795. 4989
/locus_tag="LMOf2365_0569"

CDS
3796. 4989
/locus_tag="LMOf2365_0569"
/notes="identified by similarity to OMNI:NTL01BS01695;
match to protein family HMM PF00144"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein, putative"
/protein_id="AA03351.1"
/db_xref="GI:46880051"
translation="MNRERRKHKRKNMIIVSTVFLFIATSWIYEFKSKQRTV
SAPTKKATATPFPENANDEETKEPEPEKPEKTIYKNOEIDNYLQKIGFSGS
ALVVRDGTIIQKGYMYANDEKVLNPTDTLFYIGSSQKAIATALLQLEEKIKNTN
DPISKYTPDPNGKILVNMFTSGTIGVGRKLTGNTMTDQIIKAIKRGKISQPGK
WDYLDANYAVGVKLVKSGQPLATYLOENITPAGMKHTGYQKDVGGKNTSTGYK
IDNGVQPSKPLADLKLYCAGDISMTTDMYLFDKALMDKLLISASLAKKMTSGSK
STYGMFYDPGSGVSHGVVIGWDSVNSVSHGTGYVLLSNVQNHIASTGKVNNDIY
GFLNK"
complement (5030. .5950)
/locus_tag="LMOf2365_0570"
complement (5030. 5950)
/locus_tag="LMOf2365_0570"
/notes="identified by match to protein family HMM PF01497"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate-binding protein"
/protein_id="AA03352.1"
/db_xref="GI:46880052"
translation="MYQKHAAALMTVLAVLVLTACGSSDKDSSKEKESATKVTVD
TMDKRVPTPRRIVALQNVSEILGVKPVGTDTYITTYPEATKGTESVGNKXPS
LEKIANLKPDLIIISDTIKOLLENLEKAVPVYITFGDTPDKQLSNIANLLNKAEXE
KWDXYAASKEAKATKQDLANENKAAVIOFYKGEIYHDAKFDGLYSGAGTPTPD
AAKANETKATISSEAIPEYAGADRLFILMPADGNTSDVDEMLKGVWKDIPAVKKNQV
YKVDNTKWSYDSASQLYQMEDAVKQITGR"
complement (6061. .6411)
/gene="srlB"
/locus_tag="LMOf2365_0571"
complement (6061. 6411)
/gene="srlB"
/locus_tag="LMOf2365_0571"
/EC_number="2.7.1.69"
/notes="identified by similarity to SP:P05706; match to
protein family HMM PF03829"
/codon_start=1
/transl_table=11
/product="PTS system, glucitol/sorbitol-specific, IIA
component"
/protein_id="AA03353.1"
/db_xref="GI:46880053"
translation="WTISKIMEIGLVPAPFEEKIVILFGPTATNELREISVIEHFQD
VFNALTKGNLLIGEQAYTIEVGSANKLELGHISVYFRSQNEVLPGAIVVSP
BIFPTITVGDSTIQF"
complement (6430. .7416)
/gene="srlE"
/locus_tag="LMOf2365_0572"
complement (6430. 7416)
/gene="srlE"
/locus_tag="LMOf2365_0572"
/EC_number="2.7.1.69"
/notes="identified by similarity to GP:4928286; match to
protein family HMM PF03612"
/codon_start=1
/transl_table=11
/product="PTS system, glucitol/sorbitol-specific, IIBC
component"
/protein_id="AA03354.1"
/db_xref="GI:46880054"
translation="MYQSIHYNKQGGMGVGLDITPSPRTKVTVSITGGGHPVALKI
AELSGAEAVDGFQNSIPEEMNCVVIDCGTARIVGVPMKRIPTVDVLPSPSGPLAK
HITEDIYVSGVKPEDIYVRDSTPEKPFDEKEDYAKIKEDQEEASQIKLEGLFVK
FSRGIGKAGVGYQAGRSDVMLKNIITPNAFTSMIIGIINTGIGKLIANTLSPLA
GSLGNMLLIATICSLLPFLSPILGPGCAVIAQVILGSIQAIAPVYALPALFAIN
AQAGCDFIPVGLSLAEAKPKTVFGVPAILISRMITGVLAIVIAVWFSGIMY"
complement (7437. .7958)

RESULT 36
AX417038
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
SCIENTIFIC
LOCATION/Qualifiers
1. 349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/note="seq 10 splitted into-seq 4029; from 300001 to 649980-seq 4030; from 600001 to 949980-seq 4031; from 900000 to 1163020"

ORIGIN
Alignment Scores:
Pred. No.: 1.1e-19 Length: 349980
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x AX417038 (1-349980)

QY 1 MetMetGlySerLeuPheLeuAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 339538 ATGAATCAAACTATGTATACAGCTATTCTGGGATGAATCGTTCACCAACGATTATCA 339597

QY 21 ThrValSerAsnAsnAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 339598 GTGACATCAAAATAATATTGCCAATGCCAAGCAGCGGTTATATAAACAACGCTCGTT 339657

QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
DB 339658 TTCAATGATTATTCTTACCAAAACACAAATGGGATCTGTCGAGCGGACTTTATGCTGGA 339717

QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
DB 339718 ACAAACCAATGAGCTCGGTTCCGGTTCGAAATGGGCAATTTTAAACCGATTATACA 339777

QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
DB 339778 GCAGGTTCTCCGACATCAACTGCGGAGAAACAAAGATGCACATGCAAGCGCGGCTTT 339837

QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
DB 339838 TTCAATGCTGCGGATAACGCTGGTGGGAATATCGTTTACACAGACGAGCTAGCTTTGCA 339897

QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
DB 339898 GTATCTGACATACTATTATTAACACCAACCAAGGAAATACGTTATGGATATGCAACA 339957

QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
DB 339958 GATAAAATGGCAACGTTTAAACGGAAACTTGCACCAATTCAAATCCCAATTAATAGC 340017

QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
DB 340018 GCAATTCAGCGGAGCAACAAAAATGGTAGCTTAAGCGGTAAACATTCACCTTGATTGG 340077

QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
DB 340078 GCGCAAAAA---GATACGATTCTTCCGAG----- 340104

QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
DB 340105 ----- 340107

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
DB 340108 TCTGTATACGATAATGCTGTGGAAACATAAATCAAGTCAATATGAAGCTGCGACA 340167

QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
DB 340168 CCAGATGCGAGTGGCAATGTTCTCTATGATATGAATTCAAATG-----GAC 340215

QY 254 GlySerAlaLysSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 340216 GGAAGCGTTAACTCTCTCCA-----GTAACAGGAACACATTAATAC 340257

QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 340258 AACCGCAAGCGCAACTTACAAACCCCTGATGCA----- 340290

QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 340291 ---CTTAAACATTCAAATCCACTGTTTAAACGCG----- 340326

QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 340327 -----AAACAAGTCAAT 340338

QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 340339 ATG-----GGCTTAAACCTAAGTGGCTTAAACA 340365

QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
DB 340366 AACTAC-----GGTACCAACCAA 340383

QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 340384 GTATTCTCAACCACTCTGACGGAAGGCGCTGCAACTGTAAAGACTACGAGTTACC 340443

QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 340444 GATTCTGTTATATTGCGAGTGAGTTACTCAGACGGTACAGTTATCCAGTTGCCCACTT 340503

QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
DB 340504 GCGGTTCTCTACTTCTCAATGAAGACGCTTAGTCAAAATGGGGAATGGCAATATGTT 340563

QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 340564 CCAGGATTATCTTCTGGAGATGCGATTACGGGCTTGTGCGCAAAACGAGTGGCGGA 340623

QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 340624 ATTAGCGCTCTCTATTAGAGGCTCAACGTTAGATTTATCCGCGAATTCGTTAACTTA 340683

QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
DB 340684 ATGACATACCAAGTGGTTTCCAGGCAATACAAAGTATCCGTTGTCAGATGACGTG 340743

QY 494 LeuGlnLysAlaLeuGluLeu 500
DB 340744 ATGAACCAAAATGTGAACCTTG 340764

RESULT 37
AX417039
LOCUS
DEFINITION
ACCESSION

AX417039 349980 bp DNA linear PAT 02-SEP-2002
Sequence 4030 from Patent WO0228891.
AX417039

AUTHORS Kunst, P. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4033 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES Location/Qualifiers
Source 1. 349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"

/note="seq 2058, original length: 3.011.208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"

ORIGIN

Alignment Scores:
Pred. No.: 11e-19 Length: 349980
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x AX417042 (1-349980)

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 129803 ATGAATCAAACTATGTATATACAGCTATTCTGGATGAATCGCTTCCAAACAGCATATCA 129862
QY 21 ThrValSerAsnLeuAlaAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 129863 GTGACATCAAAATATATGTCATGCGATGCAACACGACGAGGTATATAAAACAAAGCGTCGT 129922
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
DB 129923 TTCAATGATTCTTACCTTACCAAAACACAAATGGATCTGTTCGACGCGACTTTATGTGCGA 129982
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
DB 129983 ACAAAACCAATGAGCTTCGGTTCGAAATGGGCAATTTAACCGATTAACA 130042
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
DB 130043 GCAGGTCTCCGACATCACTGCAGAACAAAGATGCAGCACTGCAAGCGCGGCTTT 130102
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
DB 130103 TTCATTGCTGGCGATACGCTGGTGGGAATATCGTTTACACGACGACGAGCTTGTGCA 130162
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
DB 130163 GTATCTGACATTAATTAACCAACGACCAAGAAATACGTTATGGATATGCAACA 130222
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
DB 130223 GATAAAATGCAACGTTTAAACCGAAATCTGCAACCAATTCAAATCCCAATTAATAGC 130282
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
DB 130283 GCAATTCAGCGAAGCAACAAAATGGTAGCTTAAGCGGTAACTCCACTGTATGG 130342
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSer 194
DB 130343 GCGCAAAAA--GATACGATTTCTTCGAG----- 130369
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
DB 130370 -----CTT 130372

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
DB 130373 TCTGTATACATATCTCTGGTGAACAAATAAACTTCAATCAATGAAGCTGGACA 130432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
DB 130433 CCAGATCGAGTGGCAATGTTCTCTATGATATGAATATCAAAATG-----GAC 130480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 130481 GGAACGCTTAACCTCTCCA-----GTAACAGGAACACATTAAATAC 130522
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 130523 AACGCGCAAGCGCAACTTACAAACCCCTGATGCA----- 130555
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 130556 ---CTTAAACATTCAAATCAATCCACTGTTAAACGCG----- 130591
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 130592 -----AAACAGTCAAT 130603
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 130604 ATG-----GCTTAACCTTAAGTGGCTTAACA 130630
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
DB 130631 AACTAC-----GGTACCACCA 130648
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 130649 GTATTTCTCACCACTTCTGACGAAAGCGCTCAACTGTAAAGAGCTACGCGATTACC 130708
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 130709 GATTCTGTTATATGCTGAGTATCTCAGACGCTACAGTATCCAGTTGCCCACTT 130768
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
DB 130769 GCGGTGTCTACTTCTTCCATGAAGACGGCTTAGTCAAAATGGGGAATGCGCAATATGTT 130828
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 130829 CCAGGATATCTCTGGAGATGCGAGTTTACGCGCTTGTGCGCAACGAGCTGGCGA 130888
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 130889 ATTACGCGCTCTCTATTAGAGGCTCAACCGTAGATTATCCGCGAATTCGTTAACTTA 130948
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
DB 130949 ATGACATACCAAGTGGTTTCCAAAGGCAATCAAAAGTTATCCGTTGTCAGATGACGTG 131008
QY 494 LeuGlnLysAlaLeuGluLeu 500
DB 131009 ATGAACAAATGTGAAGTTG 131029
RESULT 39
AX641666 349980 bp DNA linear PAT 21-FEB-2003
LOCUS Sequence 2856 from Patent WO0101118.
DEFINITION AX641666
ACCESSION AX641666
VERSION AX641666.1 GI:28474427
KEYWORDS Listeria monocytogenes
SOURCE Listeria monocytogenes
ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Buchrieser, C., Frangeul, L., Couve, E., Rusniok, C., Fsihi, H.,

Dehoux, P., Dussurget, O., Chetouani, F., Nedjari, H., Glaser, P., Kunst, P., Cossart, P., Daniels, J., Goebel, W., Kreft, J., Kuhn, M., Ng, E., Vasquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P., Tierrez-Martinez, A., Amend, A., Chakraborty, T., Domann, E., Hain, T., Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Baquero, F., Garcia del Portillo, F., Gomez-Lopez, N., Maduenio, E., de Pablo, B., Weiland, J., Kaerst, U., Enlian, K. D., Hauf, J., Rose, M., and Voss, H.
 Listeria monocytogenes genome, polypeptides and uses
 Patent: WO 010118-A 2856 11-APR-2001;
 INSTITUT PASTEUR (FR)

TITLE JOURNAL

FEATURES

source

Location/Qualifiers

1. 349980

/organism="Listeria monocytogenes"

/mol_type="unassigned DNA"

/db_xref="taxon:1639"

/note="Seq 1 for big: original length: 2944528-split into 10 <223>seq 0001 : From 0.000.001 to 0.349.980

length: 349.980 <223>seq 2855 : From 0.300.001 to

0.649.980 length: 349.980 <223>seq 2856 : From 0.600.001

to 0.949.980 length: 349.980 <223>seq 2857 : From

0.900.001 to 1.249.980 length: 349.980 <223>seq 2858 :

From 1.200.001 to 1.549.980 length: 349.980 <223>seq 2859

2860 : From 1.500.001 to 1.849.980 length: 349.980 <223>seq

<223>seq 2861 : From 2.100.001 to 2.449.980 length:

349.980 <223>seq 2862 : From 2.400.001 to 2.749.980

length: 349.980 <223>seq 2863 : From 2.700.001 to

2.944.528 length: 244.528"

ORIGIN

Alignment Scores:

Pred. No.: 1.1e-19 Length: 349980
 Score: 484.50 Matches: 137
 Percent Similarity: 42.21% Conservative: 77
 Best Local Similarity: 27.02% Mismatches: 188
 Query Match: 18.74% Indels: 105
 DB: 6 Gaps: 11

US-10-009-823a-1 (1-502) x AX641666 (1-349980)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 130516 ATGAATCAAACTAGTATACAGCTATTCTCGGATGAATCGGTCCCAACAGCATTATCA 130575
 Qy 21 ThrValSerAsnAniLeaAlaAsnAlaAsnThrIleGlyTyrllysGlnGlnValVal 40
 Db 130576 GTAACATCAAAATATATGCAATGCCAACACACAGGATATATAAAACAAACGCGTGT 130635
 Qy 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
 Db 130636 TTCAATGATTTTACTTTACCAAAAACAATCGGATCTGTGACGGCGACTTTATGCTGGA 130695
 Qy 57 GlnGlyProAsnGlnAlaGlyMetGlyValaGlnValGlySerValaIlePheThr 76
 Db 130696 ACAAAACCAATAGTTTCGGTTCGGTTCGAAATTTGGGCGATTTTAACTGACTATACA 130755
 Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
 Db 130756 GCAGGTTTCCCGACTTCACTGCGAGAAATAAAGATCGACACTACAAGCGCGCGCTTT 130815
 Qy 97 PheGlnVal-----ThrLeuGluAspLysValHisThrArgAlaGlyAsnPheArg 114
 Db 130816 TTCAATGCTGGTATACCGTGTGGGATATCGTTTATACAGTACGAGTACGTTTGA 130875
 Qy 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
 Db 130876 GTATCTGACATAACTATTTTAAACGCGACGAAAGAAATACGTTATGGTTACGCAACG 130935
 Qy 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 Db 130936 GACAAAATAGTACGTTTAAACGGTAAATTTACAAACCAATCCCAATCCGCTAAATAGC 130995
 Qy 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174

Db 130996 GCAATTCACGGGAGAACCAACAAAAATGTTAGCTTAAGCGGTAACTCCCACTTGTATGG 131055
 Qy 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyPheAlaLeuLeuGluSer 194
 Db 131056 GCGGAAAAA---GATACGATTCTTCCGAG--- 131082
 Qy 195 TrpLysGlyAsnGlyThrProPheIleSerThrSerAsnTySerTyAlaGlnProMet 214
 Db 131083 -----CTG 131085
 Qy 215 ArgValTyAspGlnGlnGlyAsnSerHisAspIleThrValTyPheAspGlyAla--- 233
 Db 131086 TCTGTATACGATATGCTGGCGGAAAAACATTAACCTTCAAGTCAATATGAAGAGAGCTACA 131145
 Qy 234 ProSerSerThrGlySerLysThrPheGluTyThrLeuValAlaMetAsnProSerGluAsp 253
 Db 131146 CCAGATCGAGCGGTAAATGTTTCTACGATATGAATTCAAATG-----GAC 131193
 Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 131194 GCGCAAGCATTAACCTCTCCA-----GTGACAGGACACTTAAATTAC 131235
 Qy 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 Db 131236 AATGCGCAGGCGTAACCAACGACCCAGCGCA----- 131268
 Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 131269 ---CTTAAAAATATTCAATCAATTCAACAGTAACCGC----- 131304
 Qy 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 131305 -----AAACAAGTCAAT 131316
 Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 131317 ATG-----GGCTTAAACCTAAGTGGCTTAACA 131343
 Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 131344 AACTAC-----GGACGACCAAA 131361
 Qy 374 ThrArgGlySerGlnAspGlyTyProGlnGlyAspLeuValAspValThrIleThr 393
 Db 131362 GTATTCTCACCACACTCTGATGGCAAGGCTCTCGACTGTAAAAGATTATGCGATTACC 131421
 Qy 394 SerGluGlyLysLeuGlnGlyLysTySerAsnSerGlnValValAspPheTyAsnIle 413
 Db 131422 GATTCTGCTATATTGCAAGTGAAGTATTTCAGATGGTACAGTATCCAGTTGCCCAACTC 131481
 Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTySer 433
 Db 131482 CGCGTGTCTAGCTTCTCCAAATGAAGCGGTAGTCAAAATGGGAAACGCGAATATGTT 131541
 Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyGlyLys 453
 Db 131542 CCAGGATTTATCTCTGTGTATGTCAGTATCGGGCTCTCGCCCAAAATGCCGTGGCGGA 131601
 Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 131602 ATTAGTGGTCTTCCCTAGAGAGGATCAACGAGACTTGTCCGCTGAATTCGTTAACTTA 131661
 Qy 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 Db 131662 ATGACATATCAAGTGGTTTCCCAAGGGAATACAAAGGTTATTCTGTGGCGGATGACGTG 131721
 Qy 494 LeuGlnLysAlaLeuGluLeu 500
 Db 131722 ATGAACAAATTTGGAACCTTG 131742

RESULT 40
 AB015135

LOCUS AB015135 13374 bp DNA linear BCT 18-OCT-2002
DEFINITION Shigella flexneri 2a str. 301 section 98 of 412 of the complete
Genome.
ACCESSION AB015135 AB0005674
VERSION
KEYWORDS
SOURCE
ORGANISM
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE
AUTHORS Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,
Liu, H., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Zhang, G. W.,
Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P.,
Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,
Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.
TITLE Genome sequence of Shigella flexneri 2a: insights into
pathogenicity through comparison with genomes of Escherichia coli
K12 and O157
JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)
PUBMED 12384590
REFERENCE
AUTHORS Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,
Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
China
FEATURES
source
Location/Qualifiers
1. 13374
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
111. 1685
/gene="mvin"
/locus_tag="SF1075"
111. 1685
/gene="mvin"
/locus_tag="SF1075"
/note="Residues 14 to 524 of 524 are 99 pct identical to
residues 1 to 511 of a 511 aa protein from Escherichia
coli O157:H7 ref: NP_309474.1"
/codon_start=1
/transl_table=11
/product="putative virulence factor"
/protein_id="AA042697.1"
/db_xref="GI:24051366"
/translations="MQAFCEWVIRTPMNLKSLAAVSSMTMFSRYLGFARDAIVARI
FGMATDAPFAFKPILNLRIFPAEAFSAQAFVPIABYKQKQGEDATRVFVYVSG
LITLAVVTVAGMLAAPVIMVTAPGADTADPKALTSQLKITPPTILLISLASLV
GAILTNWRESIPAFAPTLNISMIGFALPAEPVPLALAWAVTVGIIQLVYVL
PHUKTKMLVLPINFHDAGMRVQMGPAILGVSVQSLIINIFASFELASGVS
WMTYADLMFPFSGVLGVAUGTILLPSLFSFASGNHDEYNLMWGLRCFLALPS
AVALIGLSPTVLSFYQKFTADLMTQRIAYSVGLIGLIVVKVLPAGFYRQD
IKTPVKIATVLTITLQMLNLAIFGLPKHAGLSLSIGLAACNLSLYWLRQKIFTP
QPGWMAFLRLVAVLVMGSLVGLMLHIMPENSLGTMPWRLRLRLMAVVLAGIAYFAA
LAVLGFVKFEFARTV"
complement (1725.. 2141)
/gene="flgN"
/locus_tag="SF1076"
complement (1725.. 2141)
/gene="flgN"
/locus_tag="SF1076"
/note="Residues 1 to 138 of 138 are 97 pct identical to
residues 1 to 138 of a 138 aa protein from Escherichia
coli K12 ref: NP_415588.1"
/codon_start=1
/transl_table=11
/product="protein of flagellar biosynthesis"

/protein_id="AA042698.1"
/db_xref="GI:24051367"
/translations="MRLAILDQMSAVLNDLKTVMDOEOHLSMGQINGSLQWITE
QKSSLLATDYLEQRKKEPNTANSVDISQRWGEIVTKTQQLRQMQRHNGWLEGGQIE
RNQQAEMLRPHQEPFLYGANGQTSPPHRSKKISI"
complement (2146.. 2439)
/gene="flgM"
/locus_tag="SF1077"
complement (2146.. 2439)
/gene="flgM"
/locus_tag="SF1077"
/note="Residues 1 to 97 of 97 are 100 pct identical to
residues 1 to 97 of a 97 aa protein from Escherichia coli
K12 ref: NP_415589.1"
/codon_start=1
/transl_table=11
/product="anti-FlaA (anti-sigma) factor; also known as
RflB protein"
/protein_id="AA042699.1"
/db_xref="GI:24051368"
/translations="MSIDRTSPKPVSTVQPRETTDAPVTNSRAAKTASTSTSVTL
DAQAKMQPGSSDINLVEALKLAINRNGELKMDTKIADALINEAQOQLQSN"
complement (2515.. 3174)
/gene="flgA"
/locus_tag="SF1078"
complement (2515.. 3174)
/gene="flgA"
/locus_tag="SF1078"
/note="Residues 1 to 219 of 219 are 98 pct identical to
residues 1 to 219 of a 219 aa protein from Escherichia
coli K12 ref: NP_415590.1"
/codon_start=1
/transl_table=11
/product="flagellar biosynthesis; assembly of basal-body
periplasmic P ring"
/protein_id="AA042700.1"
/db_xref="GI:24051369"
/translations="MLTIKRSVAIIILFSPLTASNLTSOLHNFPSQAQAGVSDVE
VSIATPNLLPCCPELLSMNSNRLMGNVNLARCGNDKRYLYNQVQATGNYVAA
PIVKGKLEAGNVILKRGRLDLPPTLDINQLVDVSLRDLSPDFQIQTQPRQAW
RVKAGQRVNVIASGDGFSANAEGQALNAAVAQNAVRYMVSEQVVSVDADGNILIN
L"
3329.. 3745
/gene="flgB"
/locus_tag="SF1079"
3329.. 3745
/gene="flgB"
/locus_tag="SF1079"
/note="Residues 1 to 138 of 138 are 100 pct identical to
residues 1 to 138 of a 138 aa protein from Escherichia
coli K12 ref: NP_415591.1"
/codon_start=1
/transl_table=11
/product="flagellar biosynthesis, cell-proximal portion of
basal-body rod"
/protein_id="AA042701.1"
/db_xref="GI:24051370"
/translations="MLDKLDALRPFQEQALNLRQROEVLAAANIANADTPGVQARDID
FASLKKVMQGRDATSVVLTMTSTQHPAQALTPFAELQYRIPDQPSLDGNTVDM
DRETRQFDANSLQYQMSLSALSGQIKGMNVLQSGN"
4164.. 6092
/gene="flgB"
/locus_tag="SF1080"
4164.. 6092
/gene="flgB"
/locus_tag="SF1080"
/note="Residues 241 to 642 of 642 are 98 pct identical to
residues 1 to 402 of a 402 aa protein from Escherichia
coli K12 ref: NP_415594.1"
/codon_start=1
/transl_table=11
/product="flagellar biosynthesis, hook protein"
/protein_id="AA042702.1"

Db	5802	CAAAACCAACTGCTGGGCGAGATTGTACTGGCGAACTTTGCCAACACGACGAGGCTGGCA	5861
QY	426	ArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu	445
Db	5862	TCCGAGGGCGACACGCTCTGGCTGGAGCGCAATCTTCTGGCGTGGCGCTGTGGGACA	5921
QY	446	ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAsp	465
Db	5922	GCCGGACCGGAAACTTTGGCACTCTGACCAACGGTGGCGCTGGAAGCGTCCAAACGTCGAT	5981
QY	466	MetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLys	485
Db	5982	CTCAGTAAGAAGACTGTCTAATATGATCGTTGCCCGCGTAAGTCTAAGTCTAACGCCAG	6041
QY	486	SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys	501
Db	6042	ACCATCAAAACCCAGGACCCAGATCCTCAACACGCGGGTTAACTTACGC	6089

Search completed: November 1, 2004, 06:05:13
Job time : 6762 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:16 ; Search time 24 Seconds
(without alignments)
1387.151 Million cell updates/sec

Title: US-10-009-823a-1

Perfect score: 2586

Sequence: 1 MMGSFIGATGKMKTHSTGLG.....NSKSVTTADTLMQKALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 segs, 56318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/aaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/1/aaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/1/aaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/1/aaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/1/aaa/ECTUS_COMB.pep:*
- 6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	528.5	20.4	471	4	US-09-252-991A-19526
2	475.5	18.4	412	4	US-09-543-681A-8182
3	313	12.1	265	3	US-08-483-857-8
4	307	11.9	269	4	US-09-252-991A-19715
5	287	11.1	261	3	US-08-483-857-10
6	286	11.1	265	4	US-09-543-681A-8110
7	285	11.0	264	2	US-08-436-748-8
8	237.5	9.2	263	3	US-08-483-857-5
9	229.5	8.9	262	2	US-08-436-748-7
10	228.5	8.8	263	3	US-08-483-857-7
11	223.5	8.6	263	2	US-08-436-748-5
12	220	8.5	260	2	US-08-436-748-10
13	176	6.8	2736	4	US-09-252-991A-30227
14	173.5	6.7	1095	4	US-09-206-942-69
15	173.5	6.7	1536	1	US-08-308-682-2
16	173.5	6.7	1536	1	US-08-302-832-2
17	173.5	6.7	1536	2	US-08-530-198-2
18	173.5	6.7	1536	2	US-08-469-880-2
19	173.5	6.7	1536	2	US-08-728-470-2
20	173.5	6.7	1536	2	US-08-617-697-2
21	173.5	6.7	1536	3	US-08-719-641-2
22	173.5	6.7	1536	4	US-09-206-942-67
23	169.5	6.6	248	2	US-08-436-748-6
24	165	6.4	270	3	US-08-483-857-6
25	155	6.0	240	2	US-08-436-748-3
26	155	6.0	270	3	US-08-483-857-4
27	149.5	5.8	1160	4	US-09-206-942-65

28	149.5	5.8	1188	4	US-09-206-942-63
29	143	5.5	1306	4	US-09-538-092-330
30	142	5.5	1095	4	US-09-206-942-45
31	142	5.5	1101	4	US-09-206-942-43
32	140.5	5.4	529	4	US-09-381-656-1
33	137	5.3	585	4	US-09-655-479A-20
34	136.5	5.3	2383	4	US-09-492-709A-302
35	133	5.1	1833	4	US-08-621-944A-4
36	133	5.1	1833	4	US-08-945-567D-4
37	133	5.1	1992	4	US-08-621-944A-3
38	133	5.1	1992	4	US-08-945-567D-3
39	132	5.1	2123	3	US-08-968-685A-10
40	132	5.1	2314	3	US-09-268-347-49
41	131	5.1	2048	3	US-09-268-347-48
42	130	5.0	733	4	US-09-336-115C-4
43	129.5	5.0	1475	4	US-09-538-092-1160
44	128	4.9	2090	4	US-09-538-092-1081
45	127.5	4.9	2142	4	US-09-540-236-3459

ALIGNMENTS

RESULT 1

US-09-252-991A-19526

; Sequence 19526, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 19526

; LENGTH: 471

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-19526

Query Match 20.4%; Score 528.5; DB 4; Length 471;

Best Local Similarity 28.5%; Pred. No. 8.4e-39;

Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;

QY 4 SLFTGATGKMKTHSTGLGTGVNNIANANTI GYKQQVVFQDLFSQDLAIGTSGSGPNQAG 63

Db 11 SFNIGLSGIGQAASSGLNVGTGNNIAGTVGPKQSRAEFADYAAASV-LGS----GSNPQG 65

QY 64 MGAQVGSVRTIFTGAFEPGNSVTDLAIGKGKGFQVTLDEKVVHVRAGNREFTDGFLND 123

Db 66 SGVLLSDVSQMFQKGNIDSTNSVLDLAINGFPVTSNNGAISYTRAGYFNTDKQFIVD 125

QY 124 PSGLTMSRISNNPNIKKETLEPIQLD-FNDPTVAKSPARTKSTALNAVNLGSDTDKTK 182

Db 126 NNGYRLQGYAVGPNQGLQNGVVTDLKVERANQAQATSSIQOSYNLNSLTK-----P 177

QY 183 SEAMPYFALLESWKNGTPTISTNSYIAQPMRYVDQGNSHDITVYF---DGAPST-- 237

Db 178 PTVTFP-----DPSDAATYNSSSSLGYDSQGNSHNTMSQFFIKNEPDPNATPP 225

QY 238 ---GSKTFEYLV-AMNPFSEQSAASGTDAGLMSGTWTFSSNGELKNMFTFTGS--- 290

Db 226 IPENSWTKVLLDGVNPLDSPNKT-----MSFNVTFDASQMTSVRA--PDGSTG 275

QY 291 -----ATKDLNAPQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQONWAGAPASAA 344

Db 276 PGFSIDATTNVIQFSPA---TGNEPFTP---GTGWIPASD-GKTPTTYAMNGATGAASG 327

QY 416 ARFTSGDGLRRGNHYSATLDSG-GPEFGLPGTNSYKGLSVNQLSTSNVMSREMNMI 474
Db 177 VTFANSDDLKIGSNLYRESLNSGTASAAQDGGTGAIKSGFPLENSVNLDTDEFTMI 236
QY 475 IIRGFQMSKSVTTADTMLOKALELKR 502
Db 237 VAQRGFQMSKIITTSDEILOBLVNLKR 264

RESULT 4

US-09-252-991A-19715
; Sequence 19715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107195.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19715
LENGTH: 269
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19715

Query Match 11.9%; Score 307; DB 4; Length 269;

Best Local Similarity 21.6%; Pred. No. 2.5e-19; Indels 254; Gaps 11;
Matches 108; Conservative 46; Mismatches 93

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQQQVVFQDLFSQDLAIGSTGSGPN 60
Db 9 MSLAVSVKTLGSAQDMNLTTISNLANVSTGFKEDRAEFQDLVQIRPGGQSTQDS 68
QY 61 QAGMGQVGS-VRTITFQAPENSVT-----DLAIGKGFFOVTLED-KVHYTRAGNF 113
Db 69 ELPSGLQLGTGRVVGVTQKIFTPGSLQTTQTEPLDMAVNGRGFFVLLPDGTIVSYTRDGSF 128
QY 114 RFTQDGLNDPQFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173
Db 129 HLNSDQIVTSNGFAL-----EPAL----- 148
QY 174 LGDSTDKTQSEANPYFALLESMKNGTTPISTNSYSAQPMRVYDQGNSHDITVFDGA 233
Db 149 ----- 148
QY 234 PSSTGSKTFEYLVAAMPSPEDGSAASGDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 293
Db 149 -----VVFNE-----TQFT----- 158
QY 294 DLNMQPAPLVNGLPOFSANFVGAGIQLPLTDFGKSKQONMWAGAPASAAAIGTDIGKLP 353
Db 159 -----VG----- 160
QY 354 SMPIQTSSGNSTARNSSSTRYSQDGYPGQDLVDVTITSEGLKQKYSNQVDFYNI 413
Db 161 -----QDG-----TVSVTTTNAQ-----POVIG--NI 181
QY 414 PLARFTSEDLRREGNNHYSATLDSGPEFGLPGTNSYKGLSVNQLSTSNVMSREMNMI 473
Db 182 QTADFPINAGLQAGNNLPLETSSGAPQVGTGLNGLGTVAQNTLLENSVNVVELVNM 241
QY 474 IIRGFQMSKSVTTADTM 494
Db 242 IITQRAYEMNSKVISTADQML 262

RESULT 5

US-08-483-857-10
; Sequence 10, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:

APPLICANT: Chan, Voon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
TITLE OF INVENTION: CAMPYLOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-483-857-10

Query Match 11.1%; Score 287; DB 3; Length 261;

Best Local Similarity 21.2%; Pred. No. 1.5e-17; Indels 258; Gaps 10;
Matches 108; Conservative 42; Mismatches 101

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQQQVVFQDLFSQDLAIGSTGSGPN 60
Db 1 MISSLWIAKTGLDQAQTNMDVIANLANVSTNGTKRQRAVFDLLYQ--TIROPQAGSSE 58
QY 61 QAGM-GAQVGS-VRTITF-----OGAFEPGNSVTDLAIGKGFFOVTLEDKVH-YTRAG 111
Db 59 QTTLPGLQIGTGRVPVATERLHSGQNLSTQNNSKDVAIKGQGFQVWLPDGTSAITRDG 118
QY 112 NFRFTQDGLNDPQFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 171
Db 119 SFGVDGNGQLVTAGGFQV-----PAITIPA----- 144
QY 172 VNLGSDTKTQSEANPYFALLESMKNGTTPISTNSYSAQPMRVYDQGNSHDITVFD 231
Db 145 -----NALSIITGRD 154
QY 232 GAPSTGSKTFEYLVAAMPSPEDGSAASGDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 291
Db 155 GWSVT-----QQQRA----- 166
QY 292 TKDLNMQPAPLVNGLPOFSANFVGAGIQLPLTDFGKSKQONMWAGAPASAAAIGTDIGK 351
Db 167 ----- 166
QY 352 LPSMPTQTSSGNSTARNSSSTRYSQDGYPGQDLVDVTITSEGLKQKYSNQVDFY 411
Db 167 -----PVQVGQLN----- 174

QY 412 NIPLARFTSEDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVNDMSREMY 471
Db 175 ---JTFWDTGLESIGENLYETQSSGARNSTPGLNGAGLLYQGYEITSNVNVAELV 231
QY 472 NMIIQRFQWNSKSVTTADTMQLKALEL 500
Db 232 NMIIQRAYEINSKAVSTTDQMLKLTQL 260

RESULT 6

US-09-543-681A-8110
; Sequence 8110, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8110
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8110

Query Match 11.1%; Score 286; DB 4; Length 265;
Best Local Similarity 19.7%; Pred. No. 1.9e-17; Mismatches 55; Indels 98; Gaps 9;
Matches 100; Conservative 55; Mismatches 98; Indels 254; Gaps 9;
QY 1 MMSGLFICATGMKTHSTGLGTNNANANTIGYKQVQVVDLFSQDL----AIGSTGS 56
Db 6 MRSLLWIAKTGLDAQTNMDVINSNNLANVSTNGFKRQRAVFEDLLYTIRQPCAMTSEQT 65
QY 57 QGPN--QAGMGAQVGSVRTFTQCAFEPGNSVTDLAIGKGFQVLTEDKVH-YTFAGNF 113
Db 66 NAPSGLQIGTVGRFVATERLHSGNLAQTNGTRDVAIKQGFHFVQLPGDTAYTRDGSF 125
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNV 173
Db 126 QMDQNGOLVTSSGFQIVPA-----ILPETAKKVV----- 156
QY 174 LGSDTKTQSEANPYFALLEWSKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVFDGA 233
Db 157 -----PQ-----GRDGIVSVEIGS 169
QY 234 PSSTGTSTKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKMTAFTPTGSATK 293
Db 170 PA----- 171
QY 294 DLNAPAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNMWAGAPASAAAGTIDIGKLP 353
Db 172 -----PQ-----QVQQL- 178
QY 354 SYMPIQTSSGNTARNGSSSTRYSQDGYFQGDVLDVTITSEKLGKYSNSQVDFYNI 413
Db 179 -----TLTT----- 182
QY 414 PLARFTSEDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVNDMSREMYNM 473
Db 183 ----FINDSGLESVGENLYLETASSCAPTENAPGINGAGLLYQGYETSNVNVAELVNM 238
QY 474 IITQRFQWNSKSVTTADTMQLKALEL 500
Db 239 IQTQRAYEINSKAISTSDQMLKLTQL 265

RESULT 7

US-08-436-748-8
; Sequence 8, Application US/08436748

; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: CHAN, YOON LOONG
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-436-748-8

Query Match 11.0%; Score 285; DB 2; Length 264;
Best Local Similarity 20.3%; Pred. No. 2.3e-17;
Matches 103; Conservative 47; Mismatches 108; Indels 250; Gaps 8;

QY 1 MMSGLFICATGMKTHSTGLGTNNANANTIGYKQVQVVDLFSQDLAIGSTGSQ-- 58
Db 1 MSLALYSGISGKNTQKLEFVGNLANMSTVGTQRQVTFKDMVSTIAGGSAAGGIG 60
QY 59 ---PQAGMGAQVGSVRTFTQCAFEPGNSVTDLAIGKGFQVLTEDKVHYTRAGNFRF 115
Db 61 GTNSKQIGLSSSGTIDTINHTSATQTTGRTKDLAIDGDMRVDTGSGEAYTRAGNF-- 118
QY 116 TQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNVNLG 175
Db 119 ----YLANTGTLVTGGS-----YHVNNM 138
QY 176 DSTDKTQSEANPYFALLEWSKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPS 235
Db 139 GGTIKIPTDAQSF----- 151
QY 236 STGSKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKMTAFTPTGSATKDL 295
Db 152 ----- 151
QY 296 NAMQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNMWAGAPASAAAGTIDIGKLP 355
Db 152 -----SIGSD-----SK 158
QY 356 MPIQTSSGNTARNGSSSTRYSQDGYFQGDVLDVTITSEKLGKYSNSQVDFYNIPL 415
Db 159 VQI-----VDAQGKTQ-----DGGQIGI 176
QY 416 ARTPTSEDGLRREGNNHYSATLDSG-GPEFGLPGTSNYGKLSVNOLETNSVNDMSREMYNM 474
Db 177 TTFANSGLDKIGSNUPRESINSGTASEANQPADGATAALKSTYLENSVNDVDTDETEMI 236

Db 143 -----PQATDI 149
Qy 227 TVYFDGAPSTGKTFEYLVAAMPSEDSAGSAGTDSAGLLMSGTMTFSSNGELKNMTAFT 286
Db 150 TI-----SK-----153
Qy 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLDFGKISQONWAGAPASAAAIG 346
Db 154 -----SLVQV-----159
Qy 347 TDIGKLPMPPIOTSSGNSTARNSSSTRYSODGYPQGLVDVITITSEGLQKYSNSQ 406
Db 160 -----KL-----DQPPQTV-----GOIQ-----174
Qy 407 VVDFYNIPLAFSTEDGLRREGNNHYSATLDSGGPEFGLPCTSYGKLSVNLQTSNVD 466
Db 175 -----LANFLNEGGLAIGDNLFLTAASGAATLVRASRALACCSTDTTEASNVDA 226
Qy 467 SREVMNMIIOGFQMSKSVTTADTLMQALELK 501
Db 227 VSEITALITAGRAYEMSKVISTADQMLQATSQLR 261

RESULT 10

US-08-483-857-7
; Sequence 7, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-857-7

Query Match

Best Local Similarity 8.8%; Score 228.5; DB 3; Length 263;
Matches 106; Conservative 32; Mismatches 109; Indels 269; Gaps 11;

Qy 2 MGSIFGATGKTHSTGLGVSNINANTIGYKQQVVFQDLFSQDL-AIGSTGSGPN 60
Db 1 MQALRTAASGMAAQLNVEISNNIANNMTVGFKRAREFQDLLYTIERAGSQSSTDCN 60

Qy 61 -----QAGMGAQVGSVRTIETQGAPEFGNSVTDLAIGGKGFQVTL-EDKVHYTRAGNF 113
Db 61 IVPTQVQGGVKGAGSVYRITETQGTPTTDSPLDLAQGKGMFILLPSGGTATRAGNF 120
Qy 114 R-----FTQDGLNDPSTGFTLMGRSISNNPNIKKETLEPIQLDFNDPTVAKSPAKTST 166
Db 121 STNDQGIQVTEGGLVQV-GITIP-----143
Qy 167 ALNAVNLGDSITQDQSEANFYFALLESWKNGTPISTSNYSYAQPMRYVDQGNSHDI 226
Db 144 -----143
Qy 227 TVYFDGAPSTGKTFEYLVAAMPSEDSAGSAGTDSAGLLMSGTMTFSSNGELKNMTAFT 286
Db 144 -----QNATAIT 150
Qy 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLDFGKISQONWAGAPASAAAIG 346
Db 151 ISKS-----GLVQV-----159
Qy 347 TDIGKLPMPPIOTSSGNSTARNSSSTRYSODGYPQGLVDVITITSEGLQKYSNSQ 406
Db 160 -----KL-----DQPPQTV-----GOIQ-----174
Qy 407 VVDFYNIPLAFSTEDGLRREGNNHYSATLDSGGPEFGLPCTSYGKLSVNLQTSNVD 466
Db 175 -----LANFLNEGGLAIGDNLFLTAASGAATLVRASRALACCSTDTTEASNVDA 226
Qy 467 SREVMNMIIOGFQMSKSVTTADTLMQALELK 502
Db 227 VSEITALITAGRAYEMSKVISTADQMLQATSQLR 262

RESULT 11

US-08-436-748-5
; Sequence 5, Application US/08436748
; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-436-748-5

Query Match 8.6%; Score 223.5; DB 2; Length 263;
Best Local Similarity 19.3%; Pred. No. 7.5e-12;
Matches 98; Conservative 43; Mismatches 115; Indels 253; Gaps 12;
QY 1 MMSLFTGATGKTHSTGLTGVSNINANTYGVKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MMSLHTAATGVAQQTQIDVTNNIANVNTAGFKKRAEFADLMYQVMKYAGTSTSAT 60
QY 61 QAGMGAGVG-----SVRTFTQGAPE-PGNSVTDLAIGGKGFQVFTLEDKVHYTRAGNF 113
DB 61 LSPSGIEVGUGVRPHTAVKVFTEGHLKSTSTGLDWAIAENGFFQIQL----- 108
QY 114 RFTQDGLNDPGFTLMGSRISNNPNKKTLEPIQLDNDPTVAKSPAKTSTALNAVNV 173
DB 109 -----PDCTIEMG-----NLQK-----IMK----- 123
QY 174 LGSDTDKTOSEANPYFALLESKWNGCTPPISTSNYSYAQPMRVYDQGNSHDITVYFDC 233
DB 124 -----DNEGNI----- 129
QY 234 PSSTGSKTFEYLVAMNPSEDGSAAGTDSAGLMSGTMTFSSNGELKNMTAFTPTGSATK 293
DB 130 -----VNSDGYRLLPEMTI-----PEG----- 146
QY 294 DLNMQAPLVNGLPOPSANFVAGIQLPLDPLGKSKQNMWAGAPASAAAIGTDIGKLP 353
DB 147 -----ATAINVATD-GTVS 159
QY 354 SMVPIQTSNGSNSTARNGSSSTRYSQDGYPOGDLVDVTTITSEGLKQKYSNVQVDFYNI 413
DB 160 VMLP-----GEQQTQI---GQV 174
QY 414 PLARFTSEGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDMBSREMVM 473
DB 175 ELVQFINPAGLHSMGDNLYLETGASGAPVAGIAGQDGLGTIRHGFTELSNVQLVEEMTDL 234
QY 474 IILIORGFOMNSKSVTTADTMLOKALELKR 502
DB 235 ITQRAYEAGSKAITTSDMDLGVINQLKR 263

RESULT 12
US-08-436-748-10
Sequence 10, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, YOON LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
TITLE OF INVENTION: CAMPYLOBACTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 M15
TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 260 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-10
Query Match 8.5%; Score 220; DB 2; Length 260;
Best Local Similarity 19.2%; Pred. No. 1.5e-11;
Matches 98; Conservative 49; Mismatches 103; Indels 260; Gaps 12;
QY 1 MMSLFTGATGKTHSTGLTGVSNINANTYGVKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MISALWIAKSLDAAQQTNNMEVIANNLNMGNTGTRQAVFEDLLYQ--TIROFGAQSSE 58
QY 61 QAGM--GAQVGS--VRTFT-----QGAPEPGNSVTDLAIGGKGFQVFTL--EDKVHYTRA 110
DB 59 QGNPTGLIGTGKAVATERLHLSQQGLTQTNNS--KDAVAKGQGYMPVMLPSSGESA YTRD 117
QY 111 GNFRFTQDGLNDPGFTLMGSRISNNPNKKTLEPIQLDNDPTVAKSPAKTSTALNA 170
DB 118 GSGVNGNGQLVTAGGFQV-----QFAITIPA--- 144
QY 171 VVNLGSDTDKTOSEANPYFALLESKWNGCTPPISTSNYSYAQPMRVYDQGNSHDITVYF 230
DB 145 -----NALSIHGR 153
QY 231 DGAPSTGSKTFEYLVAMNPSEDGSAAGTDSAGLMSGTMTFSSNGELKNMTAFTPTGS 290
DB 154 DGVVQVT-----QCGQAA----- 166
QY 291 ATKDLNMQAPLVNGLPOPSANFVAGIQLPLDPLGKSKQNMWAGAPASAAAIGTDIG 350
DB 167 -----PVT-----VG 171
QY 351 KLPSPMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVTTITSEGLKQKYSNVQVDF 410
DB 172 QL----- 173
QY 411 YNIPARFTSEGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDMBSREM 470
DB 174 ---NLTTFMNDTGLSIGENLFIETQSSGAANESTPALNAAALLYQTVVETSNVNVASEL 230
QY 471 VNMIIORGFOMNSKSVTTADTMLOKALEL 500
DB 231 VNMIIQVQRAYEINSKAVSTTDDMLQKLTLQ 260

RESULT 13
US-09-252-991A-30227
Sequence 30227, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30227
LENGTH: 2736
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 6.8%; Score 176; DB 4; Length 2736;
Best Local Similarity 22.6%; Pred. No. 5.4e-06;
Matches 120; Conservative 56; Mismatches 212; Indels 142; Gaps 22;
QY 20 GTVSNINANTIGYKQQVQDLFQDLAIGTSGQG-----PNOAGMGAQGVSV 71
DB 1320 GTVVNAVA-----QDPA-GNTGPGQSTTVDAVAFNPVVPNSGNL 1359
QY 72 RTFTQGAPEPNSVTDLAIGKGFQVTLDEKVVHYTRAGNFRFTQDGFNDPSGFLMG 131
DB 1360 ----LNGTAEGSVTLTDGNGNFIGQTADG-----SGNWSFTPSQPLNGTVVNTA 1409
QY 132 SRISSNNPIKKELEPIQLDNDFTVAKSPAKTSTALNAVNLGSDTDKTSQSEANPYFAL 191
DB 1410 SDAAGNTSAPATT---TVDSLSFSIQVDPSPNSGVISGTADAGNTIIITDGNPNIGQV 1465
QY 192 LESWKGNG--TPPISTSNYSYAQPMRVYDQGNSHDITVYFDG-APSSGSKTFEYLAM 248
DB 1466 TADSGGWSFTPGIPLPDGTVVNVVARSNSVDSAPAVITVDGAPAAP-----VI 1516
QY 249 NPSEDGSAAGTDSAGLIM-----SGTMTFS-----SNGELKNMTAFT 286
DB 1517 DPS-NGTEISCTAAGATVILTDGNGNFIGQATADSGSNWTFPTGTPLANGTVINAVAQD 1575
QY 287 PTG-----SATKLNWQAPLNV-----GLPQFSANFV-----GAGIQPLTILDFG 327
DB 1576 PAGNTSGPASVTDAAIA-PPAPVINPSGVVSGTAAGATVILTDGNGNPIQGVTTADGS 1634
QY 328 IKSQONWAGAPASAAAGTIDIGKLPMSMP:QTSSGNSSTARNSSSTRYSQDGPQDGL 387
DB 1635 GK-----WAFTPATPLANGTVINAL-----AQDAAGNNSPTSATVDSLAAPV 1679
QY 388 VDVITSEGLKQKYSQVDFNFIPLARTSEDGLRRREGNNHYSATLDSGPFGLPG 447
DB 1680 IDP-----SNGSVIAGTAAGATVILTDG---NGNPIQVTTADGSGNWSFTFG 1724
QY 448 TSNGYKLSVNLSTSNVDMREVMNIIIQGFQWNSKSVTTADTLMQKA 497
DB 1725 T-----PLSNGTV-----VNAVAQAAGNTSGPASTTVDSVAPAA 1759
RESULT 14
US-09-206-942-69
; Sequence 69, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-69
Query Match 6.7%; Score 173.5; DB 4; Length 1095;
Best Local Similarity 22.8%; Pred. No. 2.1e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKHTSTGLTVSNINIAN-TIGYKQQVQDLFQDLAIGTSGQGNQAGMGAQGVSV 70
DB 570 IKAGVDGNSDSATNNANLTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 618
QY 71 VRTFTQGAPEPNS-----VTDLAIGKGFQVTLDEKVVHYTRAGNFRFTQDGFND 123

DB 619 DLTIGNTSADGNAKKVTFNQVKDSKISADG-HKVTILHSKVE--TSGSNNNTEDSSDNN 675
QY 124 PSGFTLMGSRISNNPIKKELEPIQLDNDFTVAKSPAKTSTALNAVNLGSDTDKTSQ 183
DB 676 -AGLTIDARKNVTVNNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAQTS 729
QY 184 EANPYFALLESMKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 730 ----ILGGEISSGCVTLTATAGLAVNIS-----GNTVTVTAN-SCALTTLA 773
QY 239 SKTFEYLWAMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSAIKDL 295
DB 774 GSTIKGTESVTTSSQSGDIGGTISGTVVEKATESLTTQSNKIKATTGEANVTSATGTI 833
QY 296 NAWQAPLNVGLPQFSANFVGAGIQLTDFGKISQONWAGAPASAAAGTIDIGKLPMS 355
DB 834 GGTISGNTV-----VTANAGDLIVNGAE-----INATEGAATLTTSSGKL--- 875
QY 356 MPIQTSSGNSSTARNSSSTRYSQDGPQDGL--VDVITSEGLK---QGK----- 401
DB 876 ----TTEASSHITSAGQVNLQAQDGSVAGSINAANVTLTGTLTTVKGSININATSGTL 931
QY 402 YSNQVDFNFIPLARTSEDGLRRREGNNHYSATLDSGPFGLPGTSNYKLSVNOLET 461
DB 932 VINAKDAELNGAALGNHTVVNATNANGSGSVIATTS---RVNITG---DLITINGL-- 982
QY 462 SNVDSREVMNIIIQGFQWNSKSVTTADTLMQKALELKR 502
DB 983 -NI-ISKNGINIVLL-KGVKIDVKYIQFIASVDEVEIAKR 1020
RESULT 15
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
US-08-038-682-2
Query Match      6.7%; Score 173.5; DB 1; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNINIAN-TIGYKQOVVFDLPSQDLAIGSTGSGPNQAGCAQVGS 70
Db 1011 IKAGVDGSDSDATNNANLTKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTITQGAPEPGNS-----VTDLAIGKGFFQVLTEDKVHYTRAGNFRFTQDGFND 123
Db 1060 DLTIGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLSKVE--TSGSNNTTSSDNN 1116
QY 124 PSQFTLMGSRISNPNKKTLEPIOLDNFDPVAKSPAKTSTALNAVNLGDSDTKTQS 183
Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITATQGS 1170
QY 184 EANPYFALLESKWNGCT-----PPISTSNVSYAQPMPRVYDQGNSHDITVYFDGAPSSG 238
Db 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVAMNPSDEGSAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSATKOL 295
Db 1215 GSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPSM 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQGD---VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDGSVAGSINAANVTLTNTGTLTVKGSNINATSGTL 1372
QY 402 YNSQVDFYNIPLARFTSEDLRRGNHNSATLDSGGPEFGLPQTSNYKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNVNATNANGSGSVIATSS---RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREVMNMIIOQGFOMNKSVTADTMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 16
US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match      6.7%; Score 173.5; DB 1; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNINIAN-TIGYKQOVVFDLPSQDLAIGSTGSGPNQAGCAQVGS 70
Db 1011 IKAGVDGSDSDATNNANLTKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTITQGAPEPGNS-----VTDLAIGKGFFQVLTEDKVHYTRAGNFRFTQDGFND 123
Db 1060 DLTIGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLSKVE--TSGSNNTTSSDNN 1116
QY 124 PSQFTLMGSRISNPNKKTLEPIOLDNFDPVAKSPAKTSTALNAVNLGDSDTKTQS 183
Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITATQGS 1170
QY 184 EANPYFALLESKWNGCT-----PPISTSNVSYAQPMPRVYDQGNSHDITVYFDGAPSSG 238
Db 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVAMNPSDEGSAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSATKOL 295
Db 1215 GSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPSM 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQGD---VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDGSVAGSINAANVTLTNTGTLTVKGSNINATSGTL 1372
QY 402 YNSQVDFYNIPLARFTSEDLRRGNHNSATLDSGGPEFGLPQTSNYKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNVNATNANGSGSVIATSS---RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREVMNMIIOQGFOMNKSVTADTMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 17
US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
```

CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/530,198
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTSTGLGTVSNNIANAN-TIGYKQQQVFDLFSQDLAIGTSGQPNQAGMGAGVGS 70
DB 1011 IKAGVDGSDSDATNNALTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTIPTQAGPEPNS-----VTDLAIGKGFQVLTEDKHVYTRACNFRFTQDGFND 123
DB 1060 DLITGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLHSHKVE--TSGSNNTEDSSDNN 1116
QY 124 PSQFTLMGSRISNNPNIKKETLEPIQLDNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
DB 1117 -AGTIDAKNVTVNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITAGTGS 1170
QY 184 EANYPFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1171 -----ILGGIESSGSVLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVMNPSBDGSAAGTDSAGLL---MSGTMTFSSNGELKNTAFTPTGSA TKDL 295
DB 1215 GSTIKGTESVTSQSQSDIGTISGGTVEVKATESLTQNSKIKATGTEANVT SATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVAGIQPLTLDGFKSQQNMWAGAPASAAIGTDIGKLP SM 355
DB 1275 GGTISGNTVN-----VTNAGDLTVNGGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPQTSGNSTARNSSSTRYSQDGPQGLL--VDVTITSEGL---OGK----- 401
DB 1317 -----TTRASHIITSKQVNLQDGSVAGSINANVTNTTGLTIVKGSNINATSGTL 1372
QY 402 YNSQVVDFFNYPIARFTSDGLRENNHYSATLDSGGPFGLPQTSYKLSVNQLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATSS---RVNITG---DLITINGL-- 1423
QY 462 SNVDMSEMNMIIOEGFQWNSKSVTTADTMLOKALELKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQFGIASVDEIVEAKR 1461

RESULT 18

US-08-469-880-2

; Sequence 2, Application US/08469880

Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTSTGLGTVSNNIANAN-TIGYKQQQVFDLFSQDLAIGTSGQPNQAGMGAGVGS 70
DB 1011 IKAGVDGSDSDATNNALTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTIPTQAGPEPNS-----VTDLAIGKGFQVLTEDKHVYTRACNFRFTQDGFND 123
DB 1060 DLITGNTNSADGTNAKVTNFQVKDSKISADG-HKVTLHSHKVE--TSGSNNTEDSSDNN 1116
QY 124 PSQFTLMGSRISNNPNIKKETLEPIQLDNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
DB 1117 -AGTIDAKNVTVNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITAGTGS 1170
QY 184 EANYPFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1171 -----ILGGIESSGSVLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVMNPSBDGSAAGTDSAGLL---MSGTMTFSSNGELKNTAFTPTGSA TKDL 295
DB 1215 GSTIKGTESVTSQSQSDIGTISGGTVEVKATESLTQNSKIKATGTEANVT SATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVAGIQPLTLDGFKSQQNMWAGAPASAAIGTDIGKLP SM 355

Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL----- 1316
QY 356 MPIQTSSGNSSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL---OQK----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTTVKGSNINATSGTL 1372
QY 402 YNSQVVDVFNIPLARFTSEGLRREGNNHYSATLDSGGPFGPLGTSNYGKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNNATNGSGSVIATSS-----RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREMVMNIIIOQGFOMNSKSVTTADTLMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 19
US-08-728-470-2
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
US-08-728-470-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTHSTGLTGVNNANAN-TIGYKQQVVFQDLFSQDLATGSTSQGPNQAGCAQVGS 70

Db 1011 IKAGVDGENSDSDATNNANLTIKTELKLT-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTFTQGAFFGNS-----VTDLAIGKGFFQVLTEDKVVHYTRAGNFRFTQDGFND 123
Db 1060 DLTIGTNSADGTNAKVTNFQVKDSKISADG-HKVTLSKVE--TSGSNNTEDSDNN 1116
QY 124 PSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSGTDKTS 183
Db 1117 -AGLTIKAKVTVNNITSHKAVSI-----SATSGETTKTGTINATTGNVEITAGTS 1170
QY 184 EANYFALLESKWNGT-----PPISTNSYSAQPMRVYDQOQNSHDITVYFDGAPSTG 238
Db 1171 ----ILGIESSGSVTLTATGALAVSNIS-----GNTVTVTAN--SGALATLA 1214
QY 239 SKTFEYLNAMPSDEGSAASGTDAGLL---MSGTMTFSSNGELKNWTAFTPTGSAKDL 295
Db 1215 GSTIKGTESVTTSSQSGDIGTISGGIVEKATLESITQSNKIKATIGGANVTSAGTI 1274
QY 296 NAWQPAFLVNLGPQFSANFVAGIQTLDLFGIKSQOQMMWAGAPASAAAIGCTDGLKPSM 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSSGNSSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL---OQK----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTTVKGSNINATSGTL 1372
QY 402 YNSQVVDVFNIPLARFTSEGLRREGNNHYSATLDSGGPFGPLGTSNYGKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNNATNGSGSVIATSS-----RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREMVMNIIIOQGFOMNSKSVTTADTLMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 20
US-08-617-697-2
; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

```
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLTVSNNIANAN-TIGYKQQVVFQDLFSDLAIGSTGCGPNQAGMGAVQGS 70
DB 1011 IKAGVDGSDSDATNANLTIKTKEKL-----TQDLNI-----SGFNKAETAKDGS 1059

QY 71 VRTFTQGAPEPGNS-----VTDLAIGKGFQVLTEDKHYVTRAGNFRFTQDGLND 123
DB 1060 DLTIGNTSADGNTAKKVTNFQVDSKISADG-HKVTLSKVE--TSGNNNTEDSSDNN 1116

QY 124 PSGETLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTS 183
DB 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAGTGS 1170

QY 184 BANPYFALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSTG 238
DB 1171 ----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214

QY 239 SKTPEYLVAMNPSEDGSAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSATKDL 295
DB 1215 GSTINGTESVTTSQSGDIGGTISGGIVKATESLTQNSKIKATTGEANVTSAITGI 1274

QY 296 NAMOPAPLVNGLPOFSANFVGAGIQPLTDFGIKSQQNMWAGAPASAAAGTIDIGKLPMS 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316

QY 356 MPIQTSSGNSTARNSSSTRYSQDGYPOGDL--VDVTITSEGL---OQK----- 401
DB 1317 ----TTBASHSITSAKQVNLSDQGSVAGSINAVNTLTGTLTVKGSINATSGTL 1372

QY 402 YSNQVVDVFNPIPLARTSEDLRREGNNHYSATLDSGGPEFGLPGTSYNGKLSVNQLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTS--RVNITG---DLITINGL-- 1423

QY 462 SNVDMREMNWMIIOQGFQWNSKSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 21
US-08-719-641-2
; Sequence 2, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
FILE REFERENCE: 1038-861 M5:jb
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 67
LENGTH: 1536
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-206-942-67

Query Match 6.7%; Score 173.5; DB 4; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MTHSTGLTGVNNANAN-TTGYKQQVVFQDLFSQDLAIGSTGSGPNQKMGQAQVGS 70
DB 1011 IRAGVDGNSDGNANLTIKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059

QY 71 VRTIFQGAPEFGNS-----VTDLAIGKGFQVLTEDKVHYTRAGNERFTQDGLND 123
DB 1060 DUTIGNTNGADGTNAKVTNFQKQSKISADG-HKVTLLHSKVE--TSGNNNTDSSDNN 1116

QY 124 PSFGTLMGSKRIINNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNVLDGSDTKTQS 183
DB 1117 -AGLTIDAKXNVVNNITSHKAVSI-----SATSGEITTKGTGTINATGNVEITAGTGS 1170

QY 184 EAPYFALLESKNGKT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSTG 238
DB 1171 ----ILGGIESSGVSLLPATGALAVSNIS-----GNTVTVTAN-SGALTILA 1214

QY 239 SKTFEYLVAMNPSEDGSAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSAIKDL 295
DB 1215 GSTIKTESVTTSSQSGDIGIISGTVVEKATESITQSNKIKATTGEANVTSATGTI 1274

QY 296 NAWQAPLVNGLPQPSANFVGAGIQPLTLDGFKSQQNMWAGAPASAAAIGTIDGKLPSPM 355
DB 1275 GGTISGNTVN-----VTNAGDLTVGNGAE-----INATEGAATLTTSSGKL--- 1316

QY 356 MPIQTSSGNSSTARNGSSSTRYSQDGYPOGDL--VDVITITSEGL---CGK----- 401
DB 1317 ----TTEASHITSAKQVNLGAQDSVAGSINAANVLTGTLTVKGSINATSGTL 1372

QY 402 YNSQVVDVFNIPFLAFTSEDLREGNNHYSATLDSGGPFGPLPQTSNYGKLSVNQLET 461
DB 1373 VINAKDAELNGAALGNHTVNVNATNGSGSVIATTS---RVNITG---DLITINGL-- 1423

QY 462 SNVDMREVMNLIIOFGQMSKSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 23
US-08-436-748-6
Sequence 6, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, YOON LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/436,748
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 M5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-6

Query Match 6.6%; Score 169.5; DB 2; Length 248;
Best Local Similarity 27.0%; Pred. No. 4.7e-07;
Matches 70; Conservative 35; Mismatches 121; Indels 33; Gaps 9;

QY 1 MMGSLFGATGKMTHTGLTGVNNANANTIGYKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MDNALYVGLSRQMTVRRELDIVANNANANTTGFKVEDLMVRT--EQAKPAKTLDGSSPV 58

QY 61 QAGMGAQVGSVRTFTQGAPEFGNSVTDLAIGKGFQVLTEDKVHYTRAGNERFTQDGF 120
DB 59 KFNVDT--GVRNFTQGPMTKGGDYDLAINGGFKVQANGERTDRGRFTTNEGI 115

QY 121 LNDPSGFTLM---GSRISNNPNKIKETLEPIQLDFNDPTVAKS-----PAKST 166
DB 116 LVTQAGAPVLDGSGGQITIDPR-----LGPVTVG-KDGIVSQGAIRVSRIGLVRPDDLST 169

QY 167 ALNAVNLGSDTKTQ-----SEANPYFALLESKNGKTPEISTSNYSYAQPMRVYDQGN 222
DB 170 FAKDGNLYRNTTNTAPQPVTDQAIHQGMLEA--SNVQPVIEITKLEIQ--RAYESVAK 225

QY 223 SHDITVYFDGAPSTGSKT 241
DB 226 MMDNTAELSRTPSSVWARS 244

RESULT 24
US-08-483-857-6
Sequence 6, Application US/08483857
Patent No. 6020125
GENERAL INFORMATION:
APPLICANT: Chan, Yoon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-857-6

Query Match 6.4%; Score 165; DB 3; Length 247;
Best Local Similarity 27.0%; Pred. No. 1.2e-06;
Matches 70; Conservative 36; Mismatches 119; Indels 34; Gaps 10;

QY 1 MMGLFICATGKMTSTGLGTGVSNNIANANTIGYKQVQVFDLFSQDLAIGSTGSGPN 60
DB 1 MDNALYVGLSRQMTVRRELDIVANNANANTGKVEDLMVRT--EQAKPAKILDGSSPV 58
QY 61 QAGMGAQGVSVRTFTQGAPEPGNSVTDLAIGKGFQVFTLEDKVHYTRAGNFRFTQDGF 120
DB 59 KFMVMT---GVRNFTQGTGTDGVDLAINGWGFQVQANGERYTRDGRFTTNPEGI 115
QY 121 LNDPSGFTLM---GSRISNNPNKKTETLEPIQLDNDPTVAKS-----PAKTTT 166
DB 116 LVTQAGAPVLDGGGQITIDPR-----LGEVTVG-KDGIVSQAIKVRISGLVRPDLST 169
QY 167 ALNAVNLGSDTKTQ----SEANPYFALLESKNGTPTTSTSNYSYAQPMRVYDQGN 222
DB 170 FAXDGNLYRNTTAPQPVTDQAIIHQGLEA--SNVQVBIETKLEIQ--RAYESVAK 225
QY 223 SHLTVYFDGAPSTGSKT 241
DB 226 MND-TAELSRTPSSVWARS 243

RESULT 25
US-08-436-748-3
Sequence 3, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, VOON LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 MIS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-3

Query Match 6.0%; Score 155; DB 2; Length 270;
Best Local Similarity 30.7%; Pred. No. 1.1e-05;
Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGLFICATGKMTSTGLGTGVSNNIANANTIGYKQVQV---FQDLFSO---DLAIGS- 53
DB 1 MONGYQATGGMVTFQFKLDVITNNLNANTSGYKRDVVIADFKRIFKETQDELPIENH 60
QY 54 --TSGQPNQAGMG-AQVGSVRTFTQGAPEPGNSVTDLAIGKGFQV--TLEDKVHYT 108
DB 61 TRDASRFVNTTIDGPVQSQBYTDFSLGSLKATNNPLDLAMTREDAFYLVQTKDGEVRLT 120
QY 109 RAGNFRFTQDGLNDPSGFTLMGSRISNNP 138
DB 121 KQGNFQDDEGLYKQGYKVLSSDYFNNP 150

RESULT 26
US-08-483-857-4
Sequence 4, Application US/08483857
Patent No. 6020125
GENERAL INFORMATION:
APPLICANT: Chan, Voon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-857-4

Query Match 6.0%; Score 155; DB 3; Length 270;
 Best Local Similarity 30.7%; Pred. No. 1.1e-05;
 Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGSLFIGATGKMTSTGLTGNINANTTGYKQQQV---FDLFSQ---DLAIGS- 53
 Db 1 MONGYQATGGWVTOFNKLDVITNNLANINTSGYKRDVVIADFKRIFKETQDELPIENH 60

QY 54 --TGGGPNQAGW-AQGVSVTITFQGAFFGNSVTDLAIGKGFQV--TLEKQVYT 108
 Db 61 TRDASRFVNTTIDGIPQVQSEYTFSLGSLKATNPLDLAMTREDAFYLVQTKDGEVRLT 120

QY 109 RAGNFRFTODGFLNDSGFTLMGSRISNNP 138
 Db 121 KQNFOLDDEGLVNVKQGYKVLSSDYFNP 150

RESULT 27
 US-09-206-942-65
 ; Sequence 65, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: 1038-861 MIS:jb
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER APPLICATION NUMBER: 09/167,568
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 65
 ; LENGTH: 1180
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-09-206-942-65

Query Match 5.8%; Score 149.5; DB 4; Length 1180;
 Best Local Similarity 20.2%; Pred. No. 0.00034;
 Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;

QY 30 NTIGYKQQVVFODLFSQDLAIGSTGSGPNQAGMGAQVGSVRTFTQGAFFGNSVTDL 89
 Db 606 NISGFKAKIVAKD--SSNLTIGNSDDSGNT-----SAKTV-----TFNNVKDS 647

QY 90 AIGKGFFQVTLKDVHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQ 149
 Db 648 KISADG-HKVTLSKVK-TLSDNNDNTEGSDNN-TGLTITAKDVEVNNIT----- 696

QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTSQ-----EANPYFALLESW--- 195
 Db 697 ---SHKTVNVSANGGITTKTGTINATAGNVEITAGTSGIOGGIESKPGSVTIIVAGGDT 753

QY 196 -----KGN-----GTPPISTSNYSYAQPMRVYDQGN----- 222
 Db 754 LAVGNISGNAVTVTANGSALTTLAGSTIKGTESITSS-----QSGNIGGKISG 802

QY 223 -----SHDITVYFDGAPSGTSGKTFEYLAVNPNPSDGSAAAG-----TDSAG 264
 Db 803 KTVNVTATSLTTQADSKIEATGEG---EANTVSKTSIIIGGTISGTVTEATEGLTQAG 859

QY 265 LLMSGTMTFSSNGELKNMTAFTPTG-----SATKDL-----NAWQAPALVN 305
 Db 860 STITGTESVTSSQSGNIGMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTID 919

QY 306 GLPQSFANFVAGIQLPTLDGFIKSGQNMWAGAPASAAAGTIDIGKLPSPWPIQTSSGNS 365
 Db 920 GTISGNTVNTANGDITVDEDAK-----IDATGGAATLTATSGKL-----TTKASS 966

QY 366 TARNGSSSTRYSQDGPQDGL--VDVTITSEGLQKYSNSQVDFYNIPLAFTSEGD 423

Db 967 SITSAANQVNLAKDSIGGINAANVTNTGALTIVKSSINANSGLVINAKDAELN 1026
 QY 424 LRREGNNHYSATLDSGGPEFGLPCTGNYKLSVNLQETSNDV-MSREVMNIIQRFQOM 482
 Db 1027 GEASGNTVNVATNANGSVIATTSRVNITGDLITINGLNIISKNGINTVILL-KGVKI 1085

QY 483 NSKSVTTADTLMLOKALEKR 502
 Db 1086 DVKIQPGIASVDEVEIAKR 1105

RESULT 28
 US-09-206-942-63
 ; Sequence 63, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: 1038-861 MIS:jb
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER APPLICATION NUMBER: 09/167,568
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 63
 ; LENGTH: 1188
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-09-206-942-63

Query Match 5.8%; Score 149.5; DB 4; Length 1188;
 Best Local Similarity 20.2%; Pred. No. 0.00034;
 Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;

QY 30 NTIGYKQQVVFODLFSQDLAIGSTGSGPNQAGMGAQVGSVRTFTQGAFFGNSVTDL 89
 Db 614 NISGFKAKIVAKD--SSNLTIGNSDDSGNT-----SAKTV-----TFNNVKDS 655

QY 90 AIGKGFFQVTLKDVHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQ 149
 Db 656 KISADG-HKVTLSKVK-TLSDNNDNTEGSDNN-TGLTITAKDVEVNNIT----- 704

QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTSQ-----EANPYFALLESW--- 195
 Db 705 ---SHKTVNVSANGGITTKTGTINATAGNVEITAGTSGIOGGIESKPGSVTIIVAGGDT 761

QY 196 -----KGN-----GTPPISTSNYSYAQPMRVYDQGN----- 222
 Db 762 LAVGNISGNAVTVTANGSALTTLAGSTIKGTESITSS-----QSGNIGGKISG 810

QY 223 -----SHDITVYFDGAPSGTSGKTFEYLAVNPNPSDGSAAAG-----TDSAG 264
 Db 811 KTVNVTATSLTTQADSKIEATGEG---EANTVSKTSIIIGGTISGTVTEATEGLTQAG 867

QY 265 LLMSGTMTFSSNGELKNMTAFTPTG-----SATKDL-----NAWQAPALVN 305
 Db 868 STITGTESVTSSQSGNIGMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTID 927

QY 306 GLPQSFANFVAGIQLPTLDGFIKSGQNMWAGAPASAAAGTIDIGKLPSPWPIQTSSGNS 365
 Db 928 GTISGNTVNTANGDITVDEDAK-----IDATGGAATLTATSGKL-----TTKASS 974

QY 366 TARNGSSSTRYSQDGPQDGL--VDVTITSEGLQKYSNSQVDFYNIPLAFTSEGD 423
 Db 975 SITSAANQVNLAKDSIGGINAANVTNTGALTIVKSSINANSGLVINAKDAELN 1034

QY 424 LRREGNNHYSATLDSGGPEFGLPCTGNYKLSVNLQETSNDV-MSREVMNIIQRFQOM 482

Db 1035 CEASGNHTVNVATNANGSGSVIATTSRVNITGDLITINGLIISKNGINTVLL-KGVKI 1093

Qy 483 NKSQVTTADTWMLOKALEKR 502

Db 1094 DVKIQPGIASVDEVEAKR 1113

RESULT 29

US-09-538-092-330
; Sequence 330, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 330
; LENGTH: 1306
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YGR014W
US-09-538-092-330

Query Match 5.5%; Score 143; DB 4; Length 1306;
Best Local Similarity 19.6%; Pred. No. 0.0015;
Matches 122; Conservative 75; Mismatches 245; Indels 182; Gaps 25;
Qy 4 SLFICATGMKTHSTGLGTGVNANNIANTIGYKQOVVFO--DLFSODLAIGS-TGSQGN 60
Db 598 SPYTAGASTEASL--ISSTSAETSOVSQSTTALQTSFASSTTEGSETSQGPS 655
Qy 61 QAGGAQVGSVRIFPTGAPGNSVTDL-----AIGGKGFQVLEKXVHYTRAG 111
Db 656 TSVVLVQMPs-----STSEFSPQTTQMNASSSSQYTISSGILSQVSDTSVSYTSS 711
Qy 112 NFRFTQDGLNDPSGFTLMGSRIS-----NNPNIKKETEPIQLDFNDPTVAK- 159
Db 712 S---SVSQVSDTPVSYTSSSVSQVSDTPVSYTSSSVSQVSDTPVSYTSSSVSQV 768
Qy 160 --SPAKTSTALNAVNLGSDTKTQSEANPYFALLESWKNGTTPPISTSNYSVA----- 211
Db 769 SDTPVSYTSSSVSQVSDTSVPSTSRSSVQV-----SDTPVPSSTSRSSVQTSSS 822
Qy 212 -QMRYVYDQ-----GRSHDI-----TVYED 231
Db 823 LQPTTSSQRTTSTHCALSESSSVSQASEITSSINATASEVHSIQTTAATQSYTLST 882
Qy 232 GAPSTSGTKTFEYLVAMPSEDSGASAGTDSAGLMSGTMTFSSNGELKMTAFTPTGSA 291
Db 883 DANSSASAPLE--VATSTPTPSKASSL-----LLTPSTSSLSQVATNTNVQTSLTTEST 936
Qy 292 T-----KDLANWQAPLVNGLPQ--PSANFVGAGIQPLTL----- 324
Db 937 TVLEPSTTSSSTFSLVTSDDNNWIPTELITQAPAASTASTVGGTQMTLPHRAIAA 996
Qy 325 -----DFGIKSOQN--MWAGAPASAAIGTDIGKLP----- 353
Db 997 TQVPEPEGYTLITIGFKKALNYEFVVEPKSSAQI--FCYLPEALNTFPKVFNTITVL 1053
Qy 354 SMPIQTSSGN-----STAENGSSSTRYSQDY-----PQGLVDVT 391
Db 1054 QIVPLQDDSLNYLVSAEVYFPTAEIEELSNLITNSSA--FYTDGWTAKSMAAWVDS 1111

Qy 392 ITSEKLOGKYSNQVDFYNIPLARFTSEGLRRRGNHYSATLDSGGPFGPFTSNY 451

Db 1112 IPLTGLLHDSNS---GGSSDSSSSNSGSGSGNSNSGVSSSSGNSYQDAGTLEY 1168

Qy 452 GKLSVNOLETNSVDMDSREMANVII 475

Db 1169 SKKSNVSTSS-KSKKXIIGLVI 1191

RESULT 30

US-09-206-942-45
; Sequence 45, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-45

Query Match 5.5%; Score 142; DB 4; Length 1095;
Best Local Similarity 20.9%; Pred. No. 0.0014;
Matches 101; Conservative 76; Mismatches 229; Indels 78; Gaps 18;
Qy 30 NTIGYKQOVVFDLFSODLAIGSTGSGPNQAGMAGVSGVTRITFQGAPEFGNSVTDL 89
Db 604 NISGFKRAEITAKE--CADLIIGN--SDNNNA-----NAKKVTF-----NQVKDS 645
Qy 90 AIGGKGFQVLEKXVHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETEPIQ 149
Db 646 KISADS-HNVTLSKVTESNGNDABSNNG---DGTSLTINAKNITVNNNITSHKVTNIT 701
Qy 150 LDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFALLESWKNGT-----PPIS 204
Db 702 ASENVT-----KAGTINATTSGEVETAKT---GDIKGVESTSGSVTLTATGEALA 751
Qy 205 TSNYSYQAPRVYDQGNSHDIYVFDGAPSSGTSKTFEYLVAMPSEDSGASAGTDSAG 264
Db 752 VSNIS-----GNTVTITANKGLTQAGS-TVSAINGVTASSQSGDISGITSGN 799
Qy 265 LL---MSGTMTFSSNGELKMTAFTPTGSAKOLINAWQAPLVNGLPQFSANFVGAGIQP 321
Db 800 TVKYSAGDLTTKSGSEIKAKTGEANTVSATGITGTSIGNAVN-----VTANTGD 850
Qy 322 LTLDFGKSOQNWAGAPASAAIGTDIGKLPSPMPTQSSNGSTARNSSSTRYSQDG 381
Db 851 LTVEDAAK-----IDATGGAATLTATSGKL-----TTKASSITSANNQVLSAKDG 897
Qy 382 YPOGDL--VDVTITSEKLOGKYSNQVDFYNIPLARFTSEGLRRRGNHYSATLDSG 439
Db 898 SIGGINAAVNTLTTCALTIVKSSINANSGLVINAKDAELNGEASGNHTVNVNATNAN 957
Qy 440 GPFGPFGTSGNYGKLSVNOLETNSVD--MSREMNIIIQGFQMNKSVTATDTMLQAL 498
Db 958 GSGSVIATTSRVNITGDLITINGLIISKNGINTVLL-KGVKIDVKYIQGFASVDEVI 1016
Qy 499 ELKR 502
Db 1017 EAKR 1020

```
RESULT 31
US-09-206-942-43
; Sequence 43, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:Jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1101
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-43

Query Match          5.5%; Score 142; DB 4; Length 1101;
Best Local Similarity 20.9%; Pred. No. 0.0014;
Matches 101; Conservative 76; Mismatches 229; Indels 78; Gaps 18;

QY 30 NTIGYKQQVVFODLFSQDLAIGSTGQGNPQAGMGAQVGSVRTIITQGAFFPGNSVTDL 89
Db 610 NISGPKAEITAKE--GADLIIGN--SDNNNA-----NAKVTPE-----NQVKDS 651
QY 90 AIGGKFFQVTLBDDKHVYTAGNFRITQDGLNDPGLMGRISNNPNIKETLEPIQ 149
Db 652 KISADS--HNVTLSKRVETSGNNDAESNNG---DGTSLTAINAKNITVNNNITSHKTVNIT 707
QY 150 LDENDPVAKSPAKTSTALNAVNLGSDTDKQSEANPYFALLESWKNGT-----PPIS 204
Db 708 ASENVT-----KAGTINATGSEVETAKT-----GDKGVESIGSVTLTAGEALA 757
QY 205 TSNSYQAQPMRVYDQGNSHDITVYFDGAPSGTGFYLVAMNPSEDSAGSAGTDSAG 264
Db 758 VSNIS-----GNTVTITANKGLTTQAGS--TVSAINGVTATSSQSGDISGTISGN 805
QY 265 LL---MSGTMTFSSNGELKNMTAFTTGSATKDLNAWQAPLVNGLPQFSANFVGAGIOP 321
Db 806 TVKVSAGDLTTSGSEIKAKTGEANVTSAITGTTISGNV-----VTANTGD 856
QY 322 LTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPMMPIOTSSGNSTARNSSSTRYSODG 381
Db 857 LTVEDAAK-----IDATGGAATLTATSGKL-----TTKASSITSANNQVNLKADG 903
QY 382 YPOGDL--VDVTITSEGLKQKYSNQVDFYNIPLARFTESEGLREGNNHYSATLDSG 439
Db 904 SIGGNINAAAVTLTTTGATTVKSGSSINANSGLTIVNAKDAELNGRASHNHTVVNATNAN 963
QY 440 GPERGLPSTNYGKLSVNOLETNSVD--MSREWNMIIOQFQOMGKSVTADTLMQLAL 498
Db 964 GSGSVIATTSRNVITGDLATINGLNIISKNGINTVLL--KEVKIDVKYIOPGIASVDEVI 1022
QY 499 ELKR 502
Db 1023 EAKR 1026

RESULT 32
US-09-381-656-1
; Sequence 1, Application US/09381656
; Patent No. 6645509
; GENERAL INFORMATION:
; APPLICANT: SERRE, Guy Bruno Rene
; APPLICANT: SIMON, Michel
; APPLICANT: WEBER-VIVAT, Marina
```

```
; TITLE OF INVENTION: POLYPEPTIDE EXPRESSED IN THE HORN Y LAYER OF EPIDERMIS
; TITLE OF INVENTION: AND USE THEREOF
; FILE REFERENCE: 016800-336
; CURRENT APPLICATION NUMBER: US/09/381,656
; CURRENT FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: FR 97/03899
; PRIOR FILING DATE: 1997-03-28
; PRIOR APPLICATION NUMBER: FR 97/11317
; PRIOR FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Human
US-09-381-656-1

Query Match          5.4%; Score 140.5; DB 4; Length 529;
Best Local Similarity 20.3%; Pred. No. 0.00062;
Matches 108; Conservative 60; Mismatches 182; Indels 183; Gaps 24;

QY 10 TGMKTHSTGLTVSNNTIANANTIGYKQKQVVFQDLFSQDLAIGSTGQGNPQAGMGAQV 69
Db 57 TG-KGDSGFSYSGSSSGSSSISARS-----SCGSGSSSGSSSIAQGG 101
QY 70 SVRTIFTQGAEPGNSVTDLAI-----GKGFQV-----TLEDKVHY 107
Db 102 SA-----GSPKPTGTGVSQVSYSGSSSLQAGSSQLGSSSHSGSSSHSGSSSHS 155
QY 108 TRAGNFRFTQDGLNDPGLMGRISNNPNIKETLEPIQDLDFNDPTVAKSPAKTSTA 167
Db 156 SSSSSQFQSSSSFO-----VNGSALPTNDNSVRGILNPSQ-----PGOSSSSSOTS-- 202
QY 168 LNAVNLGSDTDKQ-----SEANPYFALLESWKNGTTPPISTSNYS 209
Db 203 --GVSSSQSVSSNQRPCSDIPSPCSGGPIVSHSGPYIPSSHVSQGGQRPVV----- 255
QY 210 YQPMRVYDQGNSHDITVYFDGAPSGT-----SKTFEYLVAMNPSEDSAGSAGTDSAGLL 266
Db 256 -----VDQHGSGAPGV--QGPPCSNGLPGKPCPPIITSVDKSYGYEVVCGSSDSYL 307
QY 267 MSGTMTFSSNGELKNMTAFTTGSATKDLNAWQAPLVNGLPQF----- 310
Db 308 VPG-MTYSKG-----KIYVGYFTKE-NPVKGPSP---GVPSAAGFPPISEGYFSSNP 355
QY 311 -----SANFVGAGIOPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPMMPIOTSSGN 365
Db 356 IIPSSAASSAIAIPV-----GTGGVQLCGGSGTSGKPCSPSSSRVPSSSSISSAGSP 411
QY 366 TARNSSSTRYSQDGYPGQDLVDVTITSEGLKQKYSNQVDFYNIPLARFTESE--- 422
Db 412 YHPCGSASQSPCPPG-----TGSPSSSS-----SSQSSGK 442
QY 423 -----GLRRGNHYSATLDS-----GGPEFGLP-----GTSNYGKL 454
Db 443 ILQPCGSKSSSGHPCMSVSLTITGGPD--GSPHPDPSAKAPCGSSSAGKI 494

RESULT 33
US-09-665-479A-20
; Sequence 20, Application US/09665479A
; Patent No. 6673570
; GENERAL INFORMATION:
; APPLICANT: Itoh, Fumiko
; APPLICANT: Itoh, Susumu
; APPLICANT: Heldin, Carl-Henrik
; APPLICANT: ten-Dijke, Peter
; TITLE OF INVENTION: SMAD ASSOCIATING POLYPEPTIDES
; FILE REFERENCE: L00461.70096.US
; CURRENT APPLICATION NUMBER: US/09/665,479A
; CURRENT FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US 60/154,846
; PRIOR FILING DATE: 1999-09-20
```

```
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 585
; TYPE: PR
; ORGANISM: Xenopus laevis
US-09-665-479A-20

Query Match
Best Local Similarity 5.3%; Score 137; DB 4; Length 585;
Matches 108; Conservative 62; Mismatches 191; Indels 158; Gaps 23;

QY 10 TGMKTHSTGLTGVNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGGNQAGMGAQVG 69
Db 122 TGETNRATAPGTAANGSPAA-----PDLISGPGLSGLNGLMG-----160
QY 70 SVRIIFTQGAPEPGNSVTDLAIGKGFQVLTEDKVVHYTRAGNFRFTQDGFNDPSGFTL 129
Db 161 -----SNFWELOQOQORQLMSNPMLSQIMENPLVQNMMSNPDLMRQMIIANPQMQL 214
QY 130 MG-----SRISNNPNIKKETEPIQLDFNDPTVAKSPAKTSTAL-----NAVNLGD-----176
Db 215 MERNPEISHMLNPELMKQWE-----LARNPAMQEMMRNQDRALSNLLESIPGG 264
QY 177 -----STDKTQSEANPYFALLESWKNGTTPPISTN-----YSYAQP 213
Db 265 YNALRRMYTDIOEPMFSAAREQFGNPFSAAGSEGSASOPLRTENREPLPNPWPSPASP 324
QY 214 MRVVDQGNSHDITVYFDGAPSSGSKTFEYL-VAMNPSDGSAAAGT-----DSAGLL 266
Db 325 SS-QNQTNSN-----SNTGSTTSQSVPTVSNPLGNAASLGTGYNSPENGQLL 373
QY 267 MSGTNTSSNGELKNMTAFPTGSGATKDLNAMQP-----APLVNGLPQFSAN-----313
Db 374 QQ-----ITENPOLIQSMISAPYTRSMQMAONPEFTAQWGNIPFSGNPQLQOLRHQ 429
QY 314 --FVGAGIQP-----LTDFGKSKQNMWAGAPASAAAGTGTGKLPKM 355
Db 430 LPVFLQMQNPESVMNSNPFAMQALLOVQGLQTLQTEAPGLLSLGSVG-----IPGV 484
QY 356 MPITSGNSTARNSSSTRY-----SDGYPOGLVDVVTITSEGLQKYSNSQVV 408
Db 485 PP--TSGGSTAPENPASSSTPSSASPSSGSSNNPQQMMQMIQ---LLAG--GNSQVQ 537
QY 409 DFYNIPLARTSE-----DGLRREGNNHYSATLDSGG 440
Db 538 N-----PEVRFQSDQLDNAMGFNREAN--VQALITGG 570

RESULT 34
US-09-492-709A-302
; Sequence 302, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PR
; ORGANISM: E. Coli

Query Match
Best Local Similarity 5.3%; Score 136.5; DB 4; Length 2383;
Matches 113; Conservative 79; Mismatches 217; Indels 131; Gaps 25;

QY 16 STGLGTGVNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGGNQAGMGAQVGSVRT-- 73
Db 1505 ATVADETEGNAIANT-----EVTF--TLPEDEVKANFTLSG-----GKVIIDA 1544
QY 74 -----IFQGAPEPGNSVTDLAIGK-----GFOVLTEDKVVHYTRAGNFRFTQDGFNL 122
Db 1545 EGKAVTLTKGTAGAHVTASMTGKSEQLVNVFIADTLTAQV-----NLNVTFEDNFA 1598
QY 123 DPSGFTLMSGRISN-NPNIKKETLEPIQLDFNDPTVAKSPAKTSTAL-----NAVNLG 175
Db 1599 NNVGWTRLQATVTDGNGN-----PLA--NEAVTFLPADVSASFTLGGGSAITDIN 1648
QY 176 DSTDKTQSEANPYFALLESWKNGTTPPISTNYSYAQPMRVYDQGNSHDITVYFDGAPS 235
Db 1649 GKAEVTLSGTK-----SGTYPVTVSVNNG-----VSDTKQVTLIADAGTA 1689
QY 236 STGSKTFEYLAMNPSDGS--AASGDSAGLLMSG-----TMTFSS-----NGEL 279
Db 1690 KLASLTSYVSFVSTTEGATWTASVTDANGNPVEGKKNFRGTSVTLSSTSVETDDRGFA 1749
QY 280 KNMTAFPTGSGATKDLN-AWQAPLVNGLPQFSANFVAGIQLTL-----DFGIK 329
Db 1750 ELVLTSTEVGLKTVSASLADKPTFVIGRLNASADVNSATITSLIPEGVQVWVAQDVAVK 1809
QY 330 SQQNMWAGAPASAAAGTDTICKLPVMPLOTSSGNT-----ARNGSSSTRYSQ 379
Db 1810 AHVNDQFNPVHQPVTFSAPSSQMIISQNTVGTQGVAEVMTPTERNGSYMKASLP 1869
QY 380 DG--YPOGLDLD--VTITSEGLQKYSNSQVDFYNIPLARTSEDLRREGN-NHYS 433
Db 1870 NGASLEKLEAIDEKLTJTASSPLIGVYAPTGA-----LTATLTSANGTPEVQVINF 1924
QY 434 ATLSGGBEPGLPGTSYKLSVNOLETNSVDMREVMNMIIOGFMQMSKSVTTADTM 493
Db 1925 VT-----PE--GATLGGKVRTNSGQAPVVLTSNKVGTVYTTASPH--NGVTIQTQTTV 1975

RESULT 35
US-08-621-944A-4
; Sequence 4, Application US/08621944A
; Patent No. 6440425
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; MEMBRANE PROTEIN OF MORAXELLA
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 10
; ADDRESS: Sim & McBurney
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,944A
; FILING DATE: 26-MAR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,370
```

; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24, 973
; REFERENCE/DOCKET NUMBER: 1038-587
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 593-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-621-944A-4

Query Match 5.1%; Score 133; DB 4; Length 1833;
Best Local Similarity 20.0%; Pred. No. 0.021;
Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;

QY 7 IGATGKTHSTGLGVSNVNIANTIGYKQOVVFDLPSQDLAIGSTGSGQ-----PN 60
DB 506 VGANGIKFTNVNGSNPGTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 551
QY 61 QAGMGAQGVSVRTITQGAPEPGNSVTDL-----AIGGKGFQV-----TLEDKVHYTRAG 111
DB 552 LDQDKLQGVNVK-ITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAA 610
QY 112 NFRTOQDFLND--PSGFTLMGSRISNPP-----138
DB 611 -----INDILNTGFNL-----KNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTS 658
QY 139 -----NIKETLEPIQLDFN-----DPTVAKSPAKTSTALN-----AVNLDGS 177
DB 659 KVVYDVNVDDTTHLTGTDNKKLGKVTTKLNTKSANGNTATNFVNSDEDAVNAKDI 718
QY 178 TDKTQSEANPYFALLESKWKGNGTPPISTSNYSVAQPMRVYDQGNSHD---ITVYFDGAP 234
DB 719 AENLNTLAKE-----IHTTKGTADTALQFT-----VKKVDENNADANAITVQKQAN 768
QY 235 SSTGSKTFEYLAVNPFSD--GSAASGTDGAGLLMSGTMTFSSNG--ELKMTAFTPTGSA 291
DB 769 NQVNTLTGKENGGLNKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSIKN-----PTGSE 823
QY 292 TKDLNMQPAPLVNGLPQFSAN---FVGAGIQPLT-----323
DB 824 QIQVGA-----DGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTGSLDKSKPHLSK 876
QY 324 --LDFGKISQONWAGAPASAAAIGTDIGKLPSPMPITQSSGNSTARNSSSTRYSQDG 381
DB 877 DGINAGGKKTNIQSGEIAQNSHDAVVGKIDYDKTELENKISSAKTAKQNSLHFSV-A 935
QY 382 YPQGDVVDVTTISGKLGKQKYSNVQVDFYNIPLAFRTSBDGLRRE-----427
DB 936 DEQGN--NFTVSNP---YSSYDTSKTSQD-----VITPAGENGITTKVNGVVRVGDQTK 985
QY 428 -----GNHYSATLDSGGPEFGLPQTSN-----YKLSVNOLETSNV---DMS 467
DB 986 GLTTPKLTGVNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNLIKDEKDT 1045
QY 468 REMVNMIILIQGFQWMSKS-----VTTADTM 493
DB 1046 RAASIVDLVSAGFNLOQNGEAVDFVSTYDTV 1076

RESULT 36
US-08-945-567D-4
; Sequence 4, Application US/08945567D
; Patent No. 6448386
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.

; APPLICANT: LOOMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
; TITLE OF INVENTION: MORAXELLA
; FILE REFERENCE: 1038-745 MIS
; CURRENT APPLICATION NUMBER: US/08/945,567D
; PRIOR FILING DATE: 1996-04-29
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1833
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-08-945-567D-4

Query Match 5.1%; Score 133; DB 4; Length 1833;
Best Local Similarity 20.0%; Pred. No. 0.021;
Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;

QY 7 IGATGKTHSTGLGVSNVNIANTIGYKQOVVFDLPSQDLAIGSTGSGQ-----PN 60
DB 506 VGANGIKFTNVNGSNPGTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 551
QY 61 QAGMGAQGVSVRTITQGAPEPGNSVTDL-----AIGGKGFQV-----TLEDKVHYTRAG 111
DB 552 LDQDKLQGVNVK-ITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAA 610
QY 112 NFRTOQDFLND--PSGFTLMGSRISNPP-----138
DB 611 -----INDILNTGFNL-----KNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTS 658
QY 139 -----NIKETLEPIQLDFN-----DPTVAKSPAKTSTALN-----AVNLDGS 177
DB 659 KVVYDVNVDDTTHLTGTDNKKLGKVTTKLNTKSANGNTATNFVNSDEDAVNAKDI 718
QY 178 TDKTQSEANPYFALLESKWKGNGTPPISTSNYSVAQPMRVYDQGNSHD---ITVYFDGAP 234
DB 719 AENLNTLAKE-----IHTTKGTADTALQFT-----VKKVDENNADANAITVQKQAN 768
QY 235 SSTGSKTFEYLAVNPFSD--GSAASGTDGAGLLMSGTMTFSSNG--ELKMTAFTPTGSA 291
DB 769 NQVNTLTGKENGGLNKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSIKN-----PTGSE 823
QY 292 TKDLNMQPAPLVNGLPQFSAN---FVGAGIQPLT-----323
DB 824 QIQVGA-----DGVKFAKVNNGVVGAGIDGTTTRITRDEIGFTGTGSLDKSKPHLSK 876
QY 324 --LDFGKISQONWAGAPASAAAIGTDIGKLPSPMPITQSSGNSTARNSSSTRYSQDG 381
DB 877 DGINAGGKKTNIQSGEIAQNSHDAVVGKIDYDKTELENKISSAKTAKQNSLHFSV-A 935
QY 382 YPQGDVVDVTTISGKLGKQKYSNVQVDFYNIPLAFRTSBDGLRRE-----427
DB 936 DEQGN--NFTVSNP---YSSYDTSKTSQD-----VITPAGENGITTKVNGVVRVGDQTK 985
QY 428 -----GNHYSATLDSGGPEFGLPQTSN-----YKLSVNOLETSNV---DMS 467
DB 986 GLTTPKLTGVNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNLIKDEKDT 1045
QY 468 REMVNMIILIQGFQWMSKS-----VTTADTM 493
DB 1046 RAASIVDLVSAGFNLOQNGEAVDFVSTYDTV 1076

Db	928	NOVNTLTLLKGENGLNLIKTRDKNGTVPFINGINTTSGLKAGKSTLNDGGLSIKN-----PTGSE 982
Qy	292	TKDNLAWQPAPLVNGLPQFSAN---FVGAGIQPLT-----323
Db	983	QIUVGA-----DGVFAKYNNNGWVGAGIDGGTRITRDEIGFTGTNGSLDKSKPHLSK 1035
Qy	324	--LDFGIKSQQQMMWAGAPASAAAIGTTDIGKLPSMMPPIQTSSGNSTARNGSSSTRYSQDG 381
Db	1036	DGINAGGKKITNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHESV-A 1094
Qy	382	YPQGLVDVTLTSEGKLGCKYKSNQVDFVFNIPLARITSDEGLRRE-----427
Db	1095	DEQGN--NFTVSNP---YSSYDTSTSD-----VITFAGENGITTKVNKGWVRVGDOTK 1144
Qy	428	-----GNHYSATLDSGGPEFLPGTNSN-----YGLKSVAQLSTNV-----DMS 467
Db	1145	GLTTPKLTGVNNGKVIDSQNQNTITGLSNTLANVTNDKGSVRTTECGNIKDEDKT 1204
Qy	468	REMVMIIIOGFQWNSKS-----VTTADTM 493
Db	1205	RAASIVDVLSAGFNLCQNGEAVDFVSTYDTV 1235
 RESULT 38 US-08-945-567D-3 ; Sequence 3, Application US/08945567D ; Patent No. 6448386 ; GENERAL INFORMATION: ; APPLICANT: SASAKI, Ken ; APPLICANT: HARKNESS, Robin E. ; APPLICANT: LOOSMORE, Sheena M. ; APPLICANT: CHONG, Pele ; APPLICANT: KLEIN, Michel H. ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA ; FILE REFERENCE: 1038-745 MTS ; CURRENT APPLICATION NUMBER: US/08/945,567D ; PRIOR FILING DATE: 1996-04-29 ; PRIOR APPLICATION NUMBER: 08/431,718 ; PRIOR FILING DATE: 1995-05-01 ; PRIOR APPLICATION NUMBER: 08/478,370 ; PRIOR FILING DATE: 1995-06-07 ; PRIOR APPLICATION NUMBER: 08/621,944 ; PRIOR FILING DATE: 1996-03-26 ; PRIOR APPLICATION NUMBER: PCT/CA96/00264 ; NUMBER OF SEQ ID NOS: 10 ; SOFTWARE: Patentin Ver. 2.1 ; SEQ ID NO 3 ; LENGTH: 1992 ; TYPE: PRP ; ORGANISM: Moraxella catarrhalis US-08-945-567D-3		
 Query Match 5.1%; Score 133; DB 4; Length 1992; Best Local Similarity 20.0%; Pred. No. 0.023; Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;		
Qy	7	IGATGMKTHSTGLTVSNNIANANTIGYKQQVVFDLFSDLAIGSTGSGQ-----PN 60
Db	665	VGANGIKFTNVGNSPMTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 710
Qy	61	QAQGAQVGSVRTIFTQGAFFEGNSVTDJ-----AIGKGFFOV-----TLEDKVHYTRAG 111
Db	711	LDOQKLQGVNWK-ITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKXNAAS 769
Qy	112	NFRFTQDGLND--PSGFTLMGSRISNNP-----138
Db	770	-----INDILNTGFNL-----KNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTS 817
Qy	139	-----NIKETLEPIQLDFN-----DPVAKSPAKTSTALN-----AVNVLGDS 177
Db	818	KVVYDVNVDDTTHLTGTDNKKLGKVKTKLNKTSANGTATNFNVNSSDEDALYNARDI 877
Qy	178	TDKTSQSBANFYALLBSWKGNGTPPISTSNSYSAQPMRVYDQOGNSHD---ITVVFDCAP 234
Db	878	AENLNTLAK-----IHHTKGTADTLQTF-----VKKVDENNADDANAIVTQKXAN 927
Qy	235	SSTGSKTFEVLVANPSED--GSAAGTSOAGLLMSGTWTFSSNG-ELKNMTAFPTTCSA 291
 Query Match 5.1%; Score 133; DB 4; Length 1992; Best Local Similarity 20.0%; Pred. No. 0.023; Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;		
Qy	7	IGATGMKTHSTGLTVSNNIANANTIGYKQQVVFDLFSDLAIGSTGSGQ-----PN 60
Db	665	VGANGIKFTNVGNSPMTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 710
Qy	61	QAQGAQVGSVRTIFTQGAFFEGNSVTDJ-----AIGKGFFOV-----TLEDKVHYTRAG 111
Db	711	LDOQKLQGVNWK-ITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKXNAAS 769
Qy	112	NFRFTQDGLND--PSGFTLMGSRISNNP-----138
Db	770	-----INDILNTGFNL-----KNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTS 817
Qy	139	-----NIKETLEPIQLDFN-----DPVAKSPAKTSTALN-----AVNVLGDS 177
Db	818	KVVYDVNVDDTTHLTGTDNKKLGKVKTKLNKTSANGTATNFNVNSSDEDALYNARDI 877
Qy	178	TDKTSQSBANFYALLBSWKGNGTPPISTSNSYSAQPMRVYDQOGNSHD---ITVVFDCAP 234
Db	878	AENLNTLAK-----IHHTKGTADTLQTF-----VKKVDENNADDANAIVTQKXAN 927
Qy	235	SSTGSKTFEVLVANPSED--GSAAGTSOAGLLMSGTWTFSSNG-ELKNMTAFPTTCSA 291

[illegible]

Query Match 5.1%; Score 132; DB 3; Length 2123;
Best Local Similarity 19.9%; Pred. No. 0.032;
Matches 121; Conservative 80; Mismatches 243; Indels 164; Gaps 27;

QY 178 TDKTQSEANPYFALLBSWKGNGTPTPISTSNYSYAQPMRVYDQGNSHD---ITVYFDGAP 234
DB 878 AENLNLAXE---IHTTKGTADTAQTFT-----VKKVDENNADANAITVGQKNAN 927
QY 235 SSTGSKTFEYLVAMPSD--GSAAGTDSAGLLMSGTMTFFSNG--ELKMTAFTPTGSA 291
DB 928 NQNTLTLAGENGLNKTDKNGTMTVFGINTSLGKAGKSTLNDGGLSIKN-----PTGSE 982
QY 292 TKDNLAWQAPLVNGLPQFSAN---FVGAGIQPLT----- 323
DB 983 QIQVGA-----DGVKFAKVNNGVVGAGIDGTTITRDEIGFTGNSLDSKSPHLSK 1035
QY 324 --LDGFKSQONWAGAPASAAAGTDTIGKLPMPFIQTSSGNSTARNSSSTRYSQDG 381
DB 1036 DGINAGGKITTNIQSIEAQNDAVVGKGIYDLKTELENKISSTAKTAQNSLHEFSV-A 1094
QY 382 YPGGLDVLVTTITSEGLKQKYSQSVQVDFVNIPLARTSGEDGLRRE----- 427
DB 1095 DEQGN--NFTVSNP---YSSYDTSKTSN-----VITFAGENGITTKVNGVVRVGIDQTK 1144
QY 428 -----GNHYSATLDSGPEFGLPQTSN-----YGLKSVNQLETSNV-----DMS 467
DB 1145 GLATPKLTVGNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEXT 1204
QY 468 REMVNMIIOQGFQMSKS-----VTTADTM 493
DB 1205 RAASIVDLSAGFNQNGEAVDFVSTYDIV 1235

RESULT 39
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSLA, LAURA
; TITLE OF INVENTION: MORAXELIA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IEM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Balgwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 5.1%; Score 132; DB 3; Length 2123;
Best Local Similarity 19.9%; Pred. No. 0.032;
Matches 121; Conservative 80; Mismatches 243; Indels 164; Gaps 27;

QY 8 GATGMK--THSGTGLTVSNNTIANANTIGYKQOVVVFQDLFSQDLAIGSTGSGQPNQAGVA 66
DB 801 GVDGLKFTDNGIALDGTITTKDKVFAQD-----GSLDKSKPYLDKDKL 847
QY 67 QVGSVRTIIFTOGAPEPG-----NSVTDLAIGKGF--QVTLDEKVVHYTRAGN----- 112
DB 848 KVGVE--ITTINGINAGGKAITGLSNTLTDATNATGHTVQLGIVDSTDKTRAASIGDVLN 906
QY 113 --FRFTODG-----FLNDPS-----GFTLMGSRISNNPNIKETLEPIOLD 152
DB 907 AGFLKXNGDAKDFVSYDVTDFINGNATTAKVYDCKASKVAVDVNVDGTTIHLTGADG 966
QY 153 N-----DPTVAKSPAKTSTALNAVNLGSDSTDKTSEANPYFALLSKG-----NGT 200
DB 967 NKNQIGVKVTTTTLTKDAKDKAINFVSNVSGD--DKALINAKDIADNLNTLAGEIRNTKGT 1024
QY 201 PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGSKTFEYLVAMN-----PSEDS 255
DB 1025 ADTALQTFQVKKVXKNGDDNDADTITV---GKDAKTNOVNTLKLKXNGGLDIQTNKDG 1081
QY 256 AASGDS--AGLLMSGTMTFFSNG--ELKMTAFTPTGSAKDLNA--WQAPLVNGLPQFS 311
DB 1082 VTFGINTQSLKAGNNTTLNNGLSIKN-----TAGNEIQVQADGVKFAKVNNGV----- 1132
QY 312 ANFVGAGIQPLT-----LDGFKSQONWAGAPASAAA 344
DB 1133 ---VGAGIDGTTITRDEIGFAGTNGSLDKSPHLSKGVNAGGKITTNIQSIEAQN 1189
QY 345 ICTDIGKLPNMPIQTSSGNSTARNSSSTRYSQDGYPQGDLDVDTITSEGLKQKYSN 404
DB 1190 DAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSV-ADEQGN--NFTVSNP---YSSYD 1243
QY 405 SQVDFVNIPLARTSEDLRRE-----GNHYSATLDSGG 440
DB 1244 SKTSD-----VITFAGENGITTKVNGVVRVGIDQTKLTPKLTUVGNNGKGVIVDSQ 1298
QY 441 PEGFLPGTSN-----YGLKSVNQLETSNV-----DMSREVMNMIIOQGFQMSKS----- 486
DB 1299 GQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEXTFAASIVDLSAGFNQNGEAVD 1358
QY 487 -VTTADTM 493
DB 1359 FVSTYDIV 1366

RESULT 40
US-09-268-347-49
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-49

Query Match 5.1%; Score 132; DB 3; Length 2314;
Best Local Similarity 22.8%; Pred. No. 0.036;
Matches 113; Conservative 63; Mismatches 196; Indels 124; Gaps 26;

QY 52 GSTGSGPNQAGMGACVGSVRTITFGAFAEPGNSVTDLAIG---CKGFEQVTLF---DKV 105

1444	Db	GDGT-----TAKKLGELITI--KGQDQDTNKLTDNNINGVAGTGDGFTVVKLAKDJTNL	1494
106	Qy	HYTRAGNFRFTQDGF-----LNDP---SGFTLMGSRISNNPNIKKETEPIQLD	151
1495	Db	NSVNAGGTIDEKGISFVDANGQAKANTPVLSSANGLDLGKKRISN-----IGAAVD	1545
152	Qy	FNDETVAK---SPAKTSTANAVNLGSDTDKQTSSEANPYFALLESWKNGTGPPISTSNY	208
1546	Db	DNAVNPQNEVAKTNNLNOSNGASLPFVVYTDAN-----GXPIGTGDG	1592
209	Qy	SYAQPMEVYD---CQGNSHDITVYFGAPSGSTGKTEPYLVAMPNSBDGS---AASGTD	262
1593	Db	KPKAIGADGKYHYANANGVPVDKQKPIITDADKLANLAAGKPLDAGHQVVASLGGNS	1652
263	Qy	AGLLMGTM-----TFSSN-GLKQWTAFTPTGSATKOLNAWQAPLVNGLPOFS	311
1653	Db	DAITLTIKSLTPQIDPTPNTGNAGQAQSL-----PSSLAAQOASNAASVXDVLN-----	1702
312	Qy	ANFVGAGIQP--LTLDFGIKSQO--NNWAGAPASAAAGTDIGKLSMMPIOTS-----	361
1703	Db	---VGFNLQTNHQVDE-VKAYDTVNFVNGTGADITSVRSADGTM-SNITVNTALAATDD	1757
362	Qy	SGN--STARNGSSSTRRYSDG-YPOGDLVDVITTBEGKL-----QGYKYSNSQVV	408
1758	Db	DGNVLIAKADG---KFYKADDLMPNGLSKAGXSASDAKTPTGLSLVNPNAKGSTGDV	1813
409	Qy	DFVNIPLARE-----TSEDLGRRENNNHYSATLDSGGPEFGLPGTGSYNGKLSVQC	458
1814	Db	ALNNLSKAVPKSKDGTITTTVSSDGISIQOKNSSITLSKDGLNVGKGVISNVGK-GTKD	1872
459	Qy	LETSNVMSREMVNMI	474
1873	Db	TDAAVVOQLNEVENLL	1888

Search completed: October 26, 2004, 09:09:30
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 03:22:21 ; Search time 477 Seconds
(without alignments)
5524.548 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586
Sequence: 1 MMGSFIGATGKTHSTGLG.....NSKSVTTADTMLQKALELKR 502

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US1009823/runat_26102004_100128_1596/app_query.fasta_1.647
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US1009823@cgn_1_352/runat_26102004_100128_1596 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586	100.0	1509	4 AAC88037	Aac88037 Lawsonia
2	577	22.3	568	2 AAT69214	Aat69214 Lawsonia
3	570	22.0	569	2 AAT69204	Aat69204 Lawsonia
4	529.5	20.5	1242	8 ACA24087	Aca24087 Prokaryot
5	528.5	20.4	1416	11 ABD04351	Abd04351 Pseudomon
6	528.5	20.4	1512	11 ABD04124	Abd04124 Pseudomon

7	527.5	20.4	1389	8 ABT14574	Abt14574 Pseudomon
8	509	19.7	110000	2 AAX20248_06	Continuation (7 of
9	509	19.7	111309	2 AAX20250	Aax20250 Borrelia
10	506.5	19.6	1239	8 ACA26709	Aca26709 Prokaryot
11	491.5	19.0	1209	12 ADK13720	Adk13720 E. coli
12	484.5	18.7	110000	6 ABQ69245_07	Continuation (8 of
13	484.5	18.7	110000	6 ABQ67197_06	Continuation (7 of
14	484.5	18.7	110000	6 ABAQ3041_07	Continuation (8 of
15	483	18.7	2550	2 AAV99813	Aav99813 DNA encod
16	476	18.4	10461	2 AAX20553	Aax20553 Polynucle
17	475.5	18.4	1239	10 ADF03725	Adf03725 Bacterial
18	469.5	18.2	5998	6 ABQ71042	Abq71042 Listeria
19	448.5	17.3	1215	10 AC668483	Acf668483 Photorhab
20	448.5	17.3	102644	10 ACF65378	Acf65378 Photorhab
21	448.5	17.3	110000	10 ACF67367_12	Continuation (13 of
22	414.5	16.0	110000	6 ABAQ2787_3	Continuation (4 of
23	360.5	13.9	1912	5 AAS88896	Aas88896 DNA encod
24	358.5	13.9	110000	2 AAX20248_00	Aax20248 Borrelia
25	307	11.9	810	11 ABD04540	Abd04540 Pseudomon
26	302	11.7	1545	11 ABD04219	Abd04219 Pseudomon
27	295	11.4	783	10 ACF68481	Acf68481 Photorhab
28	291	11.3	32768	2 AAX20515	Aax20515 Polynucle
29	288	11.1	816	2 AAT67783	Aat67783 H. pylori
30	288	11.1	816	2 AAT77463	Aat77463 H. pylori
31	288	11.1	837	2 AAT68116	Aat68116 H. pylori
32	286	11.1	798	10 ADF03653	Adf03653 Bacterial
33	281.5	10.9	1800	3 AAZ88552	Aaz88552 C. jejuni
34	279	10.8	1398	5 AAS88418	Aas88418 DNA encod
35	277.5	10.7	1800	2 AAV58977	Aav58977 FlgFG ope
36	237.5	9.2	789	3 AAZ88554	Aaz88554 C. jejuni
37	237	9.2	555	6 ABK74597	Abk74597 Bacillus
38	221	8.5	742	12 ADM91627	Adm91627 Lawsonia
39	215.5	8.3	1524	6 ABK74555	Abk74555 Bacillus
40	201	7.8	1432	12 ADG32119	Adg32119 DNA encod
41	193	7.5	5445	4 AAI97965	Aai97965 Lawsonia
42	193	7.5	5445	9 ACA932323	ACA932323 Lawsonia
43	193	7.5	5445	10 ADG33882	Adg33882 L. intrac
44	193	7.5	5445	10 ADJ66787	Adj66787 Lawsonia
45	176	6.8	3129	11 ABD15269	Abd15269 Pseudomon

ALIGNMENTS

RESULT 1

AAC88037
ID AAC88037 standard; DNA; 1509 BP.
XX
AC AAC88037;
XX
DT 07-MAR-2001 (first entry)
XX
DE Lawsonia intracellularis flagellar hook protein FlgE gene SEQ ID NO:2.
XX
KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;
XX
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine; pig;
XX
KW infection; detection; identification; ds.
XX
OS Lawsonia intracellularis.
XX
PN WO200069904-A1.
XX
PD 23-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-AU000437.
XX
PR 13-MAY-1999; 99US-0133973P.
XX
PA (PTIZ) PIZER PROD INC.
XX
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
XX
PA (PIGR-) PTG RES & DEV CORP.
XX
PI Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;
PI Ankenbauer RG;

XX WPI: 2001-016210/02.
 DR P-PSDB; AAB36552.
 XX
 PT New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,
 PT useful in vaccines and diagnosis of Lawsonia infections, particularly in
 PT swine.
 XX
 PS Claim 32; Page 90-94; 97pp; English.
 XX
 CC The present sequence encodes the Lawsonia intracellularis flagellar hook
 CC protein FlgE. The present invention describes an isolated or recombinant
 CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or T-
 CC cell epitope of a FlgE (flagellar hook) polypeptide from a Lawsonia spp.
 CC (I) has antibacterial activity, and induces a specific humoral immune
 CC response. (I) are used as antigens in vaccines to prevent or treat
 CC infection by Lawsonia, in birds and animals, especially pigs, to raise
 CC specific antibodies (Ab) and to detect past or present infection. Ab are
 CC also useful in diagnosis, to detect L. intracellularis or immunologically
 CC cross-reactive species, also for identification of epitopes in FlgE.
 CC Vectors that contain nucleic acids (II) encoding (I) are also useful in
 CC genetic vaccines, and fragments of (II) are useful as primers or probes
 CC for detecting L. intracellularis or related microorganisms, in
 CC hybridisation or amplification assays
 XX
 SQ Sequence 1509 BP; 511 A; 264 C; 316 G; 418 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,88e-199 Length: 1509
 Score: 2586.00 Matches: 502
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-009-823A-1 (1-502) x AAC88037 (1-1509)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 1 ATGATGGGGAGTTGTTTATGTGTGCAACAGGTATGAAACCCATAGTACAGGGTGGGT 60
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 61 ACTGTCCTCCATATATGCTACCGCAATACCATGGGTATAGCAGCAGCAGGTAGTG 120
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 121 TTTCAAGACCTGTTTATGTCAAGATTTAGCAATAGGTCTTACTGGAGTCAAGGGGCAAC 180
 QY 61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPhe 80
 DB 181 CAGGCTGGTATGGGAGCAGGTTGGAGTGTTCGACAAATTTTACACAGGCTCTTT 240
 QY 81 GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnValThr 100
 DB 241 GAACCTGGCAATAGTGTAAACAGATCTTGTCTATGGTGGAAAAGGTTTTTTTCAGGTTACA 300
 QY 101 LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe 120
 DB 301 TTAGAGATTAAGTACATATACACAGCAGGGAATTTTCGTTTACTCAGATGGTCTTT 360
 QY 121 LeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIle 140
 DB 361 TTAATGATCCTAGCGAATTTACTTTAATGGGCTCAAGAAATATCTAATAATCCTAACATA 420
 QY 141 LysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer 160
 DB 421 AAAAGGAAACCCCTTGAACCAATTCAGTTAGTCTTTAATGATCCTACAGTACCAAGTCT 480
 QY 161 ProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLys 180
 DB 481 CCTGCAAAACAGTACAGCATTAACCGCTGTGGTAAACCTTTGGTGTAGTAGTACAGATAA 540

QY 181 ThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr 200
 DB 541 ACACAAAGTGAAGCTAATCCATCTTTCGACTCTTGGAGAGCTGGAAAGGAAATGAACA 600
 QY 201 ProProLysSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln 220
 DB 601 CCTCCTATTTCTACATCAAACTACTCATATGCAACCTATGAGAGTATATGATCAACAA 660
 QY 221 GlyAsnSerHisAspIleThrValTyrPheAspGlyValaProSerSerThrGlySerLys 240
 DB 661 GGAATTTCTCAGATATPACTGTATTTTATGAGGAGCACCCTCTTCAACAGGAGTAA 720
 QY 241 ThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaLysGlyThr 260
 DB 721 ACATTTCAATACCTTGTAGCTATCAATCTTCAATGAGAGTGAAGTGTGTCATCAGGA 780
 QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 781 GATAGTCAGTCTCTTAACTGTGNACTATGACATTTTCAAGTAAATGCGGAATTA 840
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaIlePro 300
 DB 841 AATATGACAGCTTTTACTCTCTCTGCTGCTGCAACAAAGATTTAAATGTCATGCAACA 900
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 DB 901 GCACCATAGTCAATGGTTTACCACAGTTTTTACGAAATTTTGTGTGCGGAATACAG 960
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 DB 961 CTTTTAATACCTAGACTTTTGAATTTAAAGGCCAACAGAAATATGTGGCAGGAGTCCAGCA 1020
 QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 1021 TCCGCTCTGCCATAGTACAGATATTTGGGAATTCCTCAATGATGCCAATACAAACA 1080
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 DB 1081 TCCAGCGGTAACTTCTACAGCAAGAAATGGATCATCTTCAACAGAAAGATATAGCCAAGAT 1140
 QY 381 GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly 400
 DB 1141 GGTATCTCTCAGGAGATCTAGTAGATGTCAATATTCCTCTGAAGGGAATTTACAAGT 1200
 QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
 DB 1201 AAGTATAGTAAATAGTCAGGTGTGTTGATTTTATATATATTCCTTTAGCAGCTTTACAAGT 1260
 QY 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
 DB 1261 GAGGATGATTAAGACGAGAAGGGAATAACCAATTTTCCGCAACACTTGACTCAGGTGGG 1320
 QY 441 ProGlnPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
 DB 1321 CCAGAGTTTGGATGGCCAGGAACATCTACTATGGAAAACCTTAGTGTGATCAACTTAG 1380
 QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
 DB 1381 ACTTCTAACCTAGACATGACAGAGAAATGGTTAATATGATTTATTTCAACGTGGTTTT 1440
 QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
 DB 1441 CAGATGAATAGTAAATCTGTTCACAGCAGACACAATGCTCAAAAAGCATTGGAATA 1500
 QY 501 LysArg 502
 DB 1501 AAGCGT 1506

RESULT 2

AAT69214/c

ID AAT69214 standard; DNA; 568 BP.

XX

AC AAT69214;

XX 20-AUG-1997 (first entry)
 XX Lawsonia intracellularis vaccine candidate DNA.
 DE Intestinal disease; porcine proliferative enteropathy; vaccine; ss.
 XX Lawsonia intracellularis.
 OS WO9720050-A1.
 XX 05-JUN-1997.
 XX 29-NOV-1996; 96WO-AU000767.
 XX 30-NOV-1995; 95AU-00006910.
 XX 30-NOV-1995; 95AU-00006911.
 XX (DARA-) DARATECH PTY LTD.
 PA (PIGR-) PIG RES & DEV CORP.
 XX Panaccio M, Hasse D;
 PI WPI; 1997-310605/28.
 XX Vaccine for treating or preventing Lawsonia intracellularis infection -
 PT especially in pigs, containing non-pathogenic form of bacterium or its
 PT components.
 XX Claim 25; Page 71; 94pp; English.
 XX DNA molecules (AAT69203-15) were isolated from an immunoscreening of a
 CC Lawsonia intracellularis library using experimental sera from vaccinated
 CC pigs. These nucleic acids, as well as isolated GroEL and GroES (AAT69201-
 CC 02) sequences, encode putative vaccine candidates (see also AAM16678-85)
 CC useful for protection of animals and birds against intestinal diseases.
 CC esp. protection of pigs against porcine proliferative enteropathy (PPE).
 CC They can also be used as genetic vaccines
 XX Sequence 568 BP; 184 A; 126 C; 72 G; 176 T; 0 U; 10 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1,42e-37 Length: 568
 Score: 577.00 Matches: 119
 Percent Similarity: 95.97% Conservative: 0
 Best Local Similarity: 95.97% Mismatches: 4
 Query Match: 22.31% Indels: 2
 DB: 2 Gaps: 0
 US-10-009-823A-1 (1-502) x AAT69214 (1-568)
 QY 1 MetMetGlySerLeuPheIleGlyAla-ThrGlyValSerHisSerThrGlyLeuG1 20
 DB 370 ATGATGGGGAGTTTGTATTGTTGTCACACAGGTATGAACCCATAGNACAGGGT-GG 312
 QY 20 yThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVa 40
 DB 311 TACTGTTCTCAATATATTGCTTAACCAATACCATTTGGGTATAGCAGCAACAGGTAGT 252
 QY 40 lPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAs 60
 DB 251 GTTTCAGACCTGTTTAGTCAAGATTAGCAATAGGTTTACTGGAAGTCAGGGGCCAAA 192
 QY 60 nGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPh 80
 DB 191 CCAGGCTGGTATGGGACAGAGTTTGAAGTGTTCGCAATTTTACACAGGGTCTTT 132
 QY 80 eGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValTh 100
 DB 131 TGAACCTGGCAATAGTGTAAACAGATCTTCTATTGTGGAAAAGGTTTTTTTCAGGTAC 72
 QY 100 rLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPh 120

DB 71 ATTAGAGTAAAGTACTACTATACAGCAGGGAATTTTCGTTTACTCAAGATCGTTT 12
 QY 120 eLeuAsnAsp 123
 DB 11 TTTAAATGAT 2
 RESULT 3
 AAT69204
 ID AAT69204 standard; DNA; 569 BP.
 XX AAT69204;
 AC AAT69204;
 XX 20-AUG-1997 (first entry)
 XX Lawsonia intracellularis vaccine candidate DNA.
 DE Intestinal disease; porcine proliferative enteropathy; vaccine;
 KW flagellar basal body rod protein; ss.
 OS Lawsonia intracellularis.
 XX Key Location/Qualifiers
 FH 209..568
 FT /*tag= a
 FT /note= "includes in-frame stop codon at 245..247"
 XX WO9720050-A1.
 XX 05-JUN-1997.
 XX 29-NOV-1996; 96WO-AU000767.
 XX 30-NOV-1995; 95AU-00006910.
 XX 30-NOV-1995; 95AU-00006911.
 XX (DARA-) DARATECH PTY LTD.
 PA (PIGR-) PIG RES & DEV CORP.
 XX Panaccio M, Hasse D;
 PI WPI; 1997-310605/28.
 DR P-PSDB; AAM16680.
 XX Vaccine for treating or preventing Lawsonia intracellularis infection -
 PT especially in pigs, containing non-pathogenic form of bacterium or its
 PT components.
 XX Claim 15; Page 51-52; 94pp; English.
 XX DNA molecules (AAT69203-15) were isolated from an immunoscreening of a
 CC Lawsonia intracellularis library using experimental sera from vaccinated
 CC pigs. These nucleic acids, as well as isolated GroEL and GroES (AAT69201-
 CC 02) sequences, encode putative vaccine candidates (see also AAM16678-85)
 CC useful for protection of animals and birds against intestinal diseases.
 CC esp. protection of pigs against porcine proliferative enteropathy (PPE).
 CC They can also be used as genetic vaccines. The polypeptide (AAM16680)
 CC encoded by the DNA molecule given in AAT69204 has sequence similarity to
 CC flagellar basal body rod protein
 XX Sequence 569 BP; 177 A; 72 C; 128 G; 189 T; 0 U; 3 Other;
 SQ
 Alignment Scores:
 Pred. No.: 5.23e-37 Length: 569
 Score: 570.00 Matches: 118
 Percent Similarity: 95.16% Conservative: 0
 Best Local Similarity: 95.16% Mismatches: 5
 Query Match: 22.04% Indels: 2
 DB: 2 Gaps: 0
 US-10-009-823A-1 (1-502) x AAT69204 (1-569)
 QY 1 MetMetGlySerLeuPheIleGlyAla-ThrGlyMetLysThrHisSerThrGlyLeuG1 20

Db 200 ATGATGGGAGTTGTTTATTGTTGTAAGCAGGTATGAAACCCATAGNACAGGNT-GG 258
 Qy 20 YThrValSerAsnAlaAsnAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVa 40
 Db 259 TACTGTCTCCAATAATATTGCTAAGCGCAATACCAATGGGTATAGCAGCAACAGGTAGT 318
 Qy 40 LpheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAs 60
 Db 319 GTTTCAGACCTGTTTGTAGTCAAGATTAGCAATAGGTTTACTGGAAGTCAGGGGCCAAA 378
 Qy 60 nGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPh 80
 Db 379 CCAGGCTGTATGGAGCACAGGTGGAGTGTTCGACAAATTTTACACAGGGTGCTTT 438
 Qy 80 eGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnValTh 100
 Db 439 TGAACCTGGCAATAGTGTAAACAGATCTGCTATTTGGTGGAAAGGTTTTTTTCAGGTTAC 498
 Qy 100 rLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPh 120
 Db 499 ATTAGAGCATTAAGTACACTATACAGCAGCGGAATTTTCGTTTACTCAAGATGCTTT 558
 Qy 120 eLeuAsnAsp 123
 Db 559 TTTAAATGAT 568

RESULT 4

ACA24087
 ID ACA24087 standard; DNA; 1242 BP.

AC ACA24087;
 XX ACA24087;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #5744.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.

XX Borrelia cepacia.

XX W0200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU20217.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 11957; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1242 BP; 266 A; 421 C; 397 G; 158 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,57e-33 Length: 1242
 Score: 529.50 Matches: 149
 Percent Similarity: 44.20% Conservative: 72
 Best Local Similarity: 29.80% Mismatches: 182
 Query Match: 20.48% Indels: 97
 DB: 13 Gaps: 13

US-10-009-823A-1 (1-502) x ACA24087 (1-1242)

Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAlaIleAla 27
 Db 16 GGCTTGAGCGGATGGCGCGTGGTTCGAGCAATCTCGAGTCATCGGCAACAACATCGCG 75
 Qy 28 AsnAlaAsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGln 47
 Db 76 AACGAAACACGGTGGCTTCAAGCAGGACGGCGCAACTTCGCCGACATGTACGGGAAT 135
 Qy 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
 Db 136 TCGTTCGCGAGTCGTCACAC-----ACGACGATCGGCATCGGCACGCGG 180
 Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
 Db 181 CTCGCGTCGTCGAGCAGAGAATTCGGCCAGGACGACATCAATTCGACGAGTCTGCGCTC 240
 Qy 88 AspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr 107
 Db 241 GACCTCGCATCAACGGCAACGGCTTCTTCCATGTCCAGCAACGCGGTGACCATGTAC 300
 Qy 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
 Db 301 TCGGCGCAGCGCAGGTTCCTCCATCGCACAGAACGGCGCATCTCGACGCGCAAGCGCGC 360
 Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro 147
 Db 361 AACCTGATGGGCTATGCGCGCGCGCGCGCGCGGTGTATCAACCGCGCAGACCGTCCCG 420
 Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
 Db 421 CTCGAGCGG-----CCGACCAACAACATCGCGCGCGCGCGCGAGC---AAG 465
 Qy 168 LeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
 Db 466 ATCACCGCGCGAGTTCACCTG---AACGCGCAGCAGCAAGGTCCGCGCCAG----- 513

Pred. No.: 3,96e-33 Length: 1512
 Score: 528.50 Matches: 151
 Percent Similarity: 45.84% Conservative: 86
 Best Local Similarity: 29.21% Mismatches: 205
 Query Match: 20.44% Indels: 75
 DB: 11 Gaps: 16

US-10-009-823a-1 (1-502) x ABD04124 (1-1512)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 DB 1503 AGTTTCAACATCGGCTGAGCGCATCCAGCGGGCTCTAGCGGCTGAACGTCAACCGC 1444
 QY 24 AsnAsnIleAlaAsnAlaThrIleGlyLysGlnGlnValValPheGlnAsp 43
 DB 1443 AACAAATCGCCCAACCGCGACCGTAGGCTTCAAGCAATCCCGCGGAGTTCGCCGAC 1384
 QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 DB 1383 GTCTACGCGCTCGGTG---CTGGGTTG-----GGCAGCAACCGCGAGGGC 1339
 QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 DB 1338 AGCGGGGTGCTGCTCTCGGACGCTCTCGCAGATGTTCAAGCAGGCAACATCGACTCGACC 1279
 QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
 DB 1278 AACAGGCTGCTGACTGCGCATCAACGCAACGCGCTTCTGTACACGCAACACGCG 1219
 QY 104 LysValHisThrArgAlaGlyAsnPheA-gPheThrGlnAspGlyPheLeuAsnAsp 123
 DB 1218 GCGATCAGCTACACCGCGCGCTACTTCAATACCGCAACGAGGATTTTCATCGTCAC 1159
 QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlu 143
 DB 1158 AACACGGCTACCGCTCGAGGCTATGCGGCGCAACGCGCAGTTGCGAACAGCGC 1099
 QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 DB 1098 GTGGTCACCGCTCAAGTCGAGCGCGCAATCAGCGCGCGCACCTCGAGCATC 1039
 QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
 DB 1038 CAGCAGCTGCTACACCTCAACTCGAGCTGAAG-----CCG 1003
 QY 183 SerGluAlaAsnProThrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
 DB 1002 CCGACCGTGACCGCTTC-----GATCGG 979
 QY 203 IleSerThrSerAsnThrSerTyraAlaGlnProMetArgValTyraAspGlnGlyAsn 222
 DB 978 TCCGACGCGCTACCTACCACTGCTCTCTGCTGGCATCTATGACAGCGAGGCAAC 919
 QY 223 SerHisAspIleThrValTyraPhe-----AspGlyAlaProSerSerThr----- 237
 DB 918 TCCACACCATGAGCGAGTCTTCATCAGACGAGCGGACCGCAATCGCAGCCCGCGG 859
 QY 238 -----GlySerLysThrPheGluTyraLeuVal---AlaMetAsnProSerGluAsp 253
 DB 858 ATTCCGGAGAACAGCTGGACCATGAAGTGCTGATCGACGGGTCAATCGCTCGATCGG 799
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 798 TCGAACCAAGACGCGC-----ATGAGCTTCAACGTCACTTC 763
 QY 274 SerSerAsnGlyLeuLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 DB 762 GACGCCAGCGCCAGATGACCTCGTTCGGGCG-----CCGACGCGCAGCACCGCGG 709
 QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
 DB 708 CCGGGCTTCAGCATCGACGCGACCAACGATGATCCAGTTTCAGTCCGCGC----- 658

QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
 DB 657 ACCGCAATCGCGCACTCC-----GGCACCGGCTGATTCGCGGGCGCTCG 610
 QY 325 AspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAa 344
 DB 609 GAC---GGCAGACCCCGCGACCTACGCTGGAATGCGGACCGGTGCGGCGGCGG 553
 QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
 DB 552 ATCTCTTCACATGCGCAAG-----ACCACCGAGTAC 520
 QY 365 SerThrAlaArgAsnGlySerSerThrArgArgTyraSerGlnAspGlyTyraProGln 384
 DB 519 TCCACCGCTTCGCGCAGACACCGCATC-----CAGGACGCGTACACCAAC 472
 QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyraSerAsn 404
 DB 471 GGTGAGCTGGCAGCGCTGGAATCGACGACACCGGGTGATCTGCGCGCGCTACACCAAT 412
 QY 405 SerGlnValValAspPheTyraAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
 DB 411 GCCCAGTCCCAAGTGCAGGCGCAGGTGTGTCGCGAACTTCGCCAATCCAGGGCGCTG 352
 QY 425 ArgArgGluGlyAsnAsnHisTyraSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
 DB 351 ACGCCGATCGCAAGACCTCTCGGTGCTGCTCGCGAGTCCGCGAGCGCGGTGCGC 292
 QY 445 LeuProGlyThrSerAsnTyraGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
 DB 291 GCGCGCGCTGGCGCACCTCGGGCGCTTGCAATCCGCGCGCTGGAAGCGTCCACGCTG 232
 QY 465 AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484
 DB 231 GACATCTCCACGAACTGGTGAACTCATGTCCACGCGCAACTACGAGGCCAATGCC 172
 QY 485 LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 DB 171 AAGACCATCCAGACCGAGGATGCGGTGACCCAGACCATCATCACTGCGC 121

RESULT 7
 ID ABT14574 standard; DNA; 1389 BP.
 AC ABT14574;
 DT 27-FEB-2003 (first entry)
 DE Pseudomonas aeruginosa biofilm formation-related gene #16.
 KW Gene; ds; biofilm formation modulation; biofilm-associated disease;
 cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
 catheter-associated infection; medical device-associated infection.
 OS Pseudomonas aeruginosa.
 FN WO200285295-A2.
 PD 31-OCT-2002.
 PF 19-APR-2002; 2002WO-US012532.
 XX 20-APR-2001; 2001US-0285190P.
 PR 24-OCT-2001; 2001US-0344142P.
 XX (IOWA) UNIV IOWA RES FOUND.
 PA (HARD) HARVARD COLLEGE.
 PI Whiteley M, Bangera MG, Lory S, Greenberg EP;
 XX WPI; 2003-075601/07.
 DR P-PSDB; ABJ18752.
 XX

PT Identifying compound capable of modulating biofilm formation by
 PT bacteria/bacterial antibiotic resistance, useful for treatment of biofilm
 PT associated disease.

XX Claim 1; Page 70; 154pp; English.

XX The invention comprises a method for identifying a compound capable of
 CC modulating biofilm formation by bacteria. The method of the invention is
 CC useful for identifying a compound capable of modulating biofilm formation
 CC by bacteria or modulating bacterial antibiotic resistance. The method of
 CC the invention is also useful for diagnosing and treating a subject
 CC (especially an immunocompromised human) that is afflicted with a biofilm-
 CC associated disease or disorder, such as: cystic fibrosis; AIDS; middle
 CC ear infections; acne; periodontal disease; catheter-associated infections
 CC ; and medical device-associated infections. The present DNA sequence
 CC represents a gene that is used in the invention

XX SQ Sequence 1389 BP; 291 A; 501 C; 396 G; 201 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.29e-33 Length: 1389
 Score: 527.50 Matches: 151
 Percent Similarity: 45.84% Conservative: 86
 Best Local Similarity: 29.21% Mismatches: 205
 Query Match: 20.40% Indels: 75
 DB: 8 Gaps: 16

US-10-009-823A-1 (1-502) x ABT14574 (1-1389)

Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 Db 4 AGTTTCAACATCGCGCTGACCGCATCCAGCGGCTCTAGCGGCTGAACGTCACCGCG 63
 Qy 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp 43
 Db 64 AACACATCGCCACCGCGCACCGTAGGTTCAAGCAATCCGCGGAGTTCCCGGAC 123
 Qy 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 Db 124 GTCTACGCCCGCTCGGTG---CTGGGTTCG-----GGCAGCAACCGCGAGGCG 168
 Qy 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 Db 169 AGCGCGTGTCTCTCGGAGCTTCGCAGATGTTCAAGCGGGCAACATCGACTCGACC 228
 Qy 84 AsnSerValThrAspLeuAlaIleGlyGlyPhePheGlnValThrLeuGluAsp 103
 Db 229 AACAGCGTGTGGACCTGGCCATCAACGGCACCGGCTTCTTCACACCAACCAACGGG 288
 Qy 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
 Db 289 GCGATAGCTACACCGCGCGGCTACTTCATACGACAGCAGGAGTTTCATCTCGAC 348
 Qy 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
 Db 349 AACACGGCTACCGCTTCGAGGCTATGCGTCCGGCCGACCGCCAGTTCGAGAACGGC 408
 Qy 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 Db 409 GTGTCTACCGACCTCAAGTTCGAGCGCGCAATCAGCGCGCCGACCTCGAGCATC 468
 Qy 163 LysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGln 182
 Db 469 CAGCAGTGTGTACACTCACTCGACGCTGAAG-----CCG 504
 Qy 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPro 202
 Db 505 CCGACCGTGCACCGGTTTC-----GATCGG 528
 Qy 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
 Db 529 TCCGACCGCGCTACCTACACTCGTCTCTTCGCTGGCATCTATCAGACCGAGGCAAC 588

Qy 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
 Db 589 TCCACACCATGACCGAGTCTTTCATCAAGACGCGCGGACCGAATGCGACCGCGCG 648
 Qy 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
 Db 649 ATTCCGGAGAACACTGGACCATGAAAGTGTGATCGACGCGCTCAATCGCGCTCGATCCG 708
 Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 709 TCGAACAGAGCGCG-----ATGAGCTTCAACGTCACCTTC 744
 Qy 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 Db 745 GACGCCAGCGCCAGATGATCTCGTTCCGGCG-----CCGACGCGCAGCAGCAGCGGG 798
 Qy 291 -----AlaThrLysAspLeuAsnAlaThrGlnProAlaProLeuVal 304
 Db 799 CCGGCGTTCAGATCGACGCGGACCAACCATGATCCAGTTCAGTCCGGCC----- 849
 Qy 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
 Db 850 ACTGCGCAATCGCGGACTCC-----GGCACCGGCTGGATTCGCGCGGCGCTCG 897
 Qy 325 AspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAla 344
 Db 898 GAC---GGCAAGACCCCGCGCACTACGCTCGAAGCGCGCAGCGGTGCGCGCAGCGGC 954
 Qy 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
 Db 955 ATCTCTTCGACATCGCGCAG-----ACCACCCAGTAC 987
 Qy 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384
 Db 988 TCCACCGGTTTCGCCAGAGCAGCAGCCGATC-----CAGGACGGCTACACCAAC 1035
 Qy 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
 Db 1036 GTTCAGTTGGCAGCGCTTGGAAATCGACACACCGGGGTGATCTCCGCCGCTACACCAAC 1095
 Qy 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
 Db 1096 GGCCAGTCCAAGTGCAGGCGCAGGTGCTGCGCAACTTCGCCAATCCACATCCAGGCGCTG 1155
 Qy 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
 Db 1156 AGCGCGATTCGCAAGACCTCTCGGTGTCAGTCTTCGGAGTCCGCGAGCGCGCGGTCGGC 1215
 Qy 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
 Db 1216 GCGCGCGCTCGGCGACCCCTGGGGCGGCTTCAATTCGCGCGGCTGGAAGCGTCCAACGTG 1275
 Qy 465 AspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSer 484
 Db 1276 GACATCTCCAACGAACTGGTGAACCTCATGTCACCGCGCAACTACACGCGCAATGCC 1335
 Qy 485 LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db 1336 AAGACCATCCAGACCGGAGATCGGTGACCCAGACCATCATCAACCTGCGC 1386

RESULT 8

AA20248_06
 Continuation (7 of 10) of AA20248 from base 60001 (Borrelia burgdorferi polynucleotide
 WP Sequence split into 10 fragments LOCUS AA20248 Accession AA20248

WP	Fragment Name	Begin	End
WP	AA20248_00	1	110000
WP	AA20248_01	100001	210000
WP	AA20248_02	200001	310000
WP	AA20248_03	300001	410000
WP	AA20248_04	400001	510000
WP	AA20248_05	500001	610000
WP	AA20248_06	600001	710000
WP	AA20248_07	700001	810000

PR 03-SEP-1997; 97US-0057483P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 XX
 XX
 PI Fraser C, White OR, Clayton R, Dougherty BA, Lathigra R;
 PI Smith HO;
 XX
 XX WPI; 1999-081217/07.
 DR
 XX
 XX
 PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention and
 PT therapy of infections, particularly Lyme disease.
 XX
 XX Claim 1; Page 738-800; 1128pp; English.
 XX
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for the
 CC detection, diagnosis, characterisation, prevention and therapy of Bb
 CC infections, e.g. Lyme disease. They can also be used for the production
 CC of biosynthetic products, e.g. enzymes. *Borrelia* belongs to a family of
 CC motile, spiral-shaped bacteria called Spirochetes. Spirochetes are
 CC pathogenic in humans and *Borrelia* causes epidemic and endemic relapsing
 CC fever, and Lyme borreliosis, more commonly known as Lyme disease
 CC
 SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 0 U; 10 Other;

Alignment Scores:
 Pred. No.: 3,21e-29 Length: 111309
 Score: 509.00 Matches: 150
 Percent Similarity: 41.62% Conservative: 71
 Best Local Similarity: 28.25% Mismatches: 190
 Query Match: 19.68% Indels: 120
 DB: 2 Gaps: 13

US-10-009-823A-1 (1-502) x AAX20250 (1-111309)

QY 1 MetMetGlySerLeuPheHleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db
 26546 ATGATGAGGTCITTAATCTGCTGTTCTGCGCCTTCAGAAATCAAAACAGATGAT 26605
 QY 21 ThrValSerAsnAlaAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValVal 40
 Db
 26606 GTTGTTCGTAACATATCGCAATGTAATACAAATGGCTTTAAAGAGGAGATTAAT 26665
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 Db
 26666 TTTCAAGATATGATATCGCAATCTATTCTCGGAGCTTCTCGCCCTACTGATGCTCGTGT 26725
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 Db
 26726 GGGACTAATCCCAAGCAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 26785
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
 Db
 26786 ACTCAAGAGAGCTTTTCAAGCACTCAAAAGCACTCTGATCTGCGATGATGATGATGAT 26845
 QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 Db
 26846 TTTTATTTTAAAGAGGTAAATTTGTTTATGATGATGATGATGATGATGATGATGAT 26905
 QY 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
 Db
 26906 GACTCTGATCGACATCTTGTAATCTCGCAAT-----GGAATGCAATTCAA 26953
 QY 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db
 26954 GGTGTGATGCGAAGATGATTAAGGTGAAAGGTTATAAATACAGCTTCTGATTTGAG 27013
 QY 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 Db
 27014 GATCTGATTATCCGATTGGAGATAAAGAGGAGCAAGTCTACGAAAAAATGTTACTTTT 27073

QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190
 Db
 27074 GCTTGTATATCTTGATAAGAGATTCCTTAATCAAGAGGTGCGATCTCGAGATATT 27133
 QY 191 LeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
 Db
 27134 GCACGCGGAACCTGGTGTTCATAATAATCA----- 27163
 QY 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
 Db
 27164 -----TTGTATGACAGTTTGGAAAT-----GTTAGTGTTCTT--- 27196
 QY 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
 Db
 27196 ----- 27196
 QY 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr 270
 Db
 27196 ----- 27196
 QY 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
 Db
 27197 -----GAGCTTAGA----- 27205
 QY 291 AlaThrLysAspLeuAsnAla-----TTPGlnProAlaProLeuValAsnGlyLeu 307
 Db
 27206 GTTGTAAAGATTTAAATACGCCCTAATTTATGGAATGCAACAGTATTAATAATGTT--- 27262
 QY 308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe--- 326
 Db
 27263 ---GAGCAAAATTCAAATTTT-----ACACTTGGGTTTGAC 27295
 QY 327 -----GlyIleLysSerGlnGlnAsnMetTTPAlaGly 337
 Db
 27296 AATGAGGAGCAATTCGGCTCTTTAAATGTTCAACACGAGGCAAAAGAGATATTCTTCAA 27355
 QY 338 AlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetPro 357
 Db
 27356 ATTCTATACATTAATGTTTGGGTGCAATGTAGTGAAGTTGGTGGAGCAAACT 27415
 QY 358 IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370
 Db
 27416 GTAAATTTGAAATTTGGGACACAGTTGGAAGTTACACTGATTAATCAATTTACTGAT 27475
 QY 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
 Db
 27476 TCTAGTAGCAAAAGGCTATTATTCAAGATGGATATGGCATGGGATATATGGAATAATTAT 27535
 QY 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
 Db
 27536 GAAATTCATCAAAATGCTGTATTAGTTGGCATTTATTCAATGGCATAAGACGAGATCTT 27595
 QY 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
 Db
 27596 GCGAAGATTCCTCTGCTCTTTTATGAATCCCGAGGACTTCGCAAAATCAGCGGACTACT 27655
 QY 431 HistySerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
 Db
 27656 AATTTTGTAGAAACACAGCAATTCAGGTCAAGTTAGATAGATAGGCGCAAACTGGACTTGTGA 27715
 QY 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
 Db
 27716 CTTGGTGATATAGATCTGCTGTTTGTAGAAATGGCAATGTTGATCTTCAGAGCAATTT 27775
 QY 471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
 Db
 27776 ACAGATATGATGATGACCCAAAGAGATTTTCAGCAATGGCAAAACCATTCACCTTCT 27835
 QY 491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db
 27836 GATCAATATTACAGAACTTGTAAAGATTGAAA 27868

ACA26709

XX ID ACA26709 standard; DNA; 1239 BP.
XX AC ACA26709;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #8366.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX OS Burkholderia mallei.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR P-PSDB; ABU22839.
XX DR WPI; 2003-029926/02.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 14579; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
XX CC prokaryotic essential genes. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1239 BP; 271 A; 424 C; 379 G; 165 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.84e-31	Length:	1239
Score:	506.50	Matches:	137
Percent Similarity:	43.34%	Conservative:	81
Best Local Similarity:	27.24%	Mismatches:	180
Query Match:	19.59%	Indels:	105
DB:	8	Gaps:	11

US-10-009-823A-1 (1-502) x ACA26709 (1-1239)

Qy	8	GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla	27
Db	16	GGTTTGAGCGGTTGGCGGGCGGTTCAGCGCACTCGACGTATCGGCACACACATCGCG	75
Qy	28	AsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln	47
Db	76	AACCGCAACACGGTGGCGTTTCAAGGGCAGCACCGCGCAGTTTCCGACATGTATCGGAAT	135
Qy	48	AspLeuAlaIleGlySerThrGlySerGlnGlnProAsnGlnAlaGlyMetGlyAlaGln	67
Db	136	TCGTCGCA-----TCGGCGTCAACAATCCGATCGGCATCGGCACGATG	180
Qy	68	ValGlySerValArgThrIlePheThrGlnGlnAlaPheGluProGlyAsnSerValThr	87
Db	181	CTGCATCGGTGCAGCAGCAGTTCAGCCAGGCGACGATCACCTCGAGCAGCGTCGCGTG	240
Qy	88	AspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr	107
Db	241	AACGTCGCGATCAACGGCAACGGCTTCTCCAGATGTCGAACAACGCGGTGACACCGTAC	300
Qy	108	ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe	127
Db	301	TCGGCGACGCGACGTTCCAGCGCGACAGACAGCGGTACATCGTCAATTCCGAGGGCGTG	360
Qy	128	ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGluThrLeuGluPro	147
Db	361	AACCTGATGGGTACGCGCGCAACGCGCGTGTATCAACCGCGCGCGCGCGCGCGCG	420
Qy	148	IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla	167
Db	421	CTGCAGCGG-----CCGACGACCAACATCGCGCCACCGCGACG---ACCAAG	465
Qy	168	LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro	187
Db	466	ATCACCGCGCGAGTTCACCTG---AACTCGCAGGACGCGGTGCGCGCG---ACGACGCGG	519
Qy	188	TyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProLysSerThrSerAsn	207
Db	520	TTC-----AACTACCGCATCGACGAC-----	543
Qy	208	TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr	227
Db	544	TACAACATACAGCAGCTCGGTGAGGTGTTCCACACGCTCGCGGTTCGCAGAACGTGAAC	603
Qy	228	ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla	247
Db	604	CTGTACTTCGTGAAGAGCGCGCAGCGCGCGCATGGGAAGCGTATCGCGCGCGCGCGCG	663
Qy	248	MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet	267
Db	664	AAGCGCGCGACCGATCTCGGCTCG-----	687
Qy	268	SerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrPro	287
Db	688	-----GTCAAGTTCAGCACCAGCGCGCACGATC	714
Qy	288	ThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeu	307
Db	715	ACGGGCGAGTCGACGCC-----	732
Qy	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly	327

[illegible]

XX	Kapur V, Gadgil M;
PI	WPI; 2004-238974/22.
XX	P-PSDB; ADK13784.
DR	
DR	
XX	New isolated and purified iron transport and metabolism polypeptides and
PT	encoding polynucleotides, useful in identifying potential targets for
PT	agents against pathogenic bacteria.
XX	
PS	Claim 31; SEQ ID NO 15; 185pp; English.
XX	
CC	The present sequence encodes an Escherichia coli iron transport and
CC	metabolism protein. Also described: (1) an isolated and purified
CC	polynucleotide comprising a nucleic acid sequence encoding an Escherichia
CC	coli iron transport and metabolism protein; and (2) an expression
CC	cassette comprising a nucleic acid sequence encoding a promoter operably
CC	linked to at least one of the polynucleotide sequences of (1). The
CC	Escherichia coli iron transport and metabolism proteins have
CC	antibacterial activity. The methods and compositions of the present
CC	invention are useful in identifying genes and proteins involved in
CC	bacterial iron transport and metabolism, and using such as potential
CC	targets for agents against pathogenic bacteria.
XX	
SQ	Sequence 1209 BP; 327 A; 320 C; 303 G; 259 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	2,89e-30 Length: 1209
Score:	491.50 Matches: 143
Percent Similarity:	43.95% Conservative: 75
Best Local Similarity:	28.83% Mismatches: 173
Query Match:	19.01% Indels: 105
DB:	12 Gaps: 15

Db 499 -----GTTACGCCATTTCAGCGCCAGCAATGCG 525
Qy 208 -----TyrSerTyAlaGlnProMetArgValTyAspGlnGlnGlyAsnSerHisAsp 225
Db 526 GATAGCTATAACAAAAAGGTTGGTACGCTTTTCGACAGTCAGGTAATGCTCATGAC 585
Qy 226 lIeThrValTyPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyLeu 245
Db 586 ATGAGCGTCTACTTT-----GTGAAGACCGGGGATATACTGGCAG----- 627
Qy 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
Db 628 GTCTACACCCAGGATAGCAGTATCAACAGCAGTTCGAAGACA----- 672
Qy 286 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 673 -----GGCAACACTCGAATTTAATGCTAATGCACTTA-----GTGGATGGTCG 720
Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
Db 721 ATGGCGAATAATATCGAACCGGCCAATTAAAGGTGCAGAACCGGCC----- 768
Qy 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAsp 325
Db 769 -----ACGTTTAGTCTGAGCTTCTCACTCC----- 795
Qy 326 PheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIle 345
Db 796 -----ATGACGCAAAAT----- 807
Qy 346 GlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSer 365
Db 808 -----ACCGGGCGCTAACCAATATT 825
Qy 366 ThrAlaArgAsnGlySerSerSerThrArgArgTySerGlnAspGlyTyProGlnGly 385
Db 826 GTGGCAACC-----ACCGAAGCGGTACAAACCGGCG 858
Qy 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTySerAsnSer 405
Db 859 GATCTGGTCAATATCAATCAATGACGAGTGTGCGCAACTATTCCCAACGAA 918
Qy 406 GlnValValAspPheTyAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
Db 919 CAACCAACTCTGGGGCAGATTGACTGGCGAACTTTCACCAACGAGAGGTCTGGCA 978
Qy 426 ArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyTyProGluPheGlyLeu 445
Db 979 TCGAAGGCGACAACTGCTGGTCTGCGACGCAATCTTCTGGGTGGCGGTGGGACA 1038
Qy 446 ProGlyThrSerAsnTyGlyLysLeuSerValAsnGlnGlnLeuThrSerAsnValAsp 465
Db 1039 GCGGAGACGGAACTTTGGCACCTTGACCAACGGTGGCGTGGAAAGCGTCCAACGTCGAT 1098
Qy 466 MetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLys 485
Db 1099 CTCAGTAAGAAGATGCTGATATGATGCTGGTCCCGAGTAACTCATGCTAACGCCGAG 1158
Qy 486 SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1159 ACCATCAAAACCCAGGACCGACGATCTCTCAACAGCTGGTTAACTTACGC 1206

RESULT 12
ABQ69245_07
Continuation (8 of 31) of ABQ69245 from base 700001 (Listeria innocua DNA sequence #694.
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245
WP Fragment Name Begin End
WP ABQ69245_00 1 110000
WP ABQ69245_01 100001 210000
WP ABQ69245_02 200001 310000
WP ABQ69245_03 300001 410000
WP ABQ69245_04 400001 510000

WP ABQ69245_05 500001 610000
WP ABQ69245_06 600001 710000
WP ABQ69245_07 700001 810000
WP ABQ69245_08 800001 910000
WP ABQ69245_09 900001 1010000
WP ABQ69245_10 1000001 1110000
WP ABQ69245_11 1100001 1210000
WP ABQ69245_12 1200001 1310000
WP ABQ69245_13 1300001 1410000
WP ABQ69245_14 1400001 1510000
WP ABQ69245_15 1500001 1610000
WP ABQ69245_16 1600001 1710000
WP ABQ69245_17 1700001 1810000
WP ABQ69245_18 1800001 1910000
WP ABQ69245_19 1900001 2010000
WP ABQ69245_20 2000001 2110000
WP ABQ69245_21 2100001 2210000
WP ABQ69245_22 2200001 2310000
WP ABQ69245_23 2300001 2410000
WP ABQ69245_24 2400001 2510000
WP ABQ69245_25 2500001 2610000
WP ABQ69245_26 2600001 2710000
WP ABQ69245_27 2700001 2810000
WP ABQ69245_28 2800001 2910000
WP ABQ69245_29 2900001 3010000
WP ABQ69245_30 3000001 3111208

Alignment Scores:
Pred. No.: 3e-27 Length: 110000
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x ABQ69245_07 (1-110000)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 29803 ATGAATCAAACTATGATACAGCTATTCTCGGATGATGATCGTTCACACAGCATTTATCA 29862
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyLysGlnGlnValVal 40
Db 29863 GTGACATCAATAATATTGCCAATGCGAACACACAGGATATAAAACCAAGCGTGT 29922
Qy 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 29923 TTCAATGATTACTTTTACCAAAACACAAATGGGATCTCTTGCAGCGGACTTTATGCTGGA 29982
Qy 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 29983 ACAACCAATGAGCTTCGGTTCGGTTCGAAATTCGGGCAATTTTAAACGATTATACA 30042
Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
Db 30043 GCAGGTCTCTCGACATCAACTGGCAGAAACAAAGATGCACACTGCAAGCGCGGCTTT 30102
Qy 97 PheGlnVal-----ThrLeuGluAspLysValHisTyThrArgAlaGlyAsnPheArg 114
Db 30103 TTCATTGCTGGCATAACGCTGGTGGGAATATCGTTTACACAGACGCGTACGTTTGA 30162
Qy 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 30163 GTATCTGACATAACTATTATAACCAACGCAACAGAAATACGTTATGGGATATGCAACA 30222
Qy 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 30223 GATAAAATGCAACCTTTTAAACGGAAACTTGCACCAATTCATATCCCATTAATAGC 30282
Qy 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 30283 GCAATTCACGCGAAGCAACAAAAATGGTACGTAAAGCGGTAAACATTCACCTTGTGG 30342

```
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
|||...|||
Db 30343 GCGGAAAAA---GATACGATTCTTCGCGAG--- 30369
QY 195 TrpLysGlyAsnGlyThrProLysThrSerAsnTyrSerTyrAlaGlnProMet 214
|||...|||
Db 30370 -----CTT 30372
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
|||...|||
Db 30373 TCTGTATACGATATGCTGTGGGAAAAATAAACTTCAAGTCAATATGAAGCTGGAC 30432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
|||...|||
Db 30433 CCAGATCGAGTGGCAATGTTCTCTAGATATGAATCAAATG---GAC 30480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
|||...|||
Db 30481 GGAAGCGTTAACTCTCCA---GTAACAGGAACTTAATTATAC 30522
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
|||...|||
Db 30523 AACGCGCAAGCGCAACTTACAAACCTGTATGCA--- 30555
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
|||...|||
Db 30556 ---CTTAAACCATTCATCAATCCACTGTGTAACGGC--- 30591
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
|||...|||
Db 30592 -----AAACAGTCAAT 30603
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
|||...|||
Db 30604 ATG-----GGCTTAACCTAAGTGCTTTAACA 30630
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
|||...|||
Db 30631 AACTAC-----GGTACCAACCAA 30648
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
|||...|||
Db 30649 GTATTCACCAACTCTGACGGAAGGCGTGCACCTGTAAAGACTACCGAGTTACC 30708
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
|||...|||
Db 30709 GATTCGTGTATATTGACGTAGTACTCAGCGGTACAGTTATCCAGTTCGCCCACTT 30768
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
|||...|||
Db 30769 GCGGTGTCTACTTCTCCCAATGAAGCGCTTAGTCAAAATGGGAATGGCGAATATGTT 30828
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
|||...|||
Db 30829 CCAGGATATCTCTGAGATGCAAGTTACGCGGTGTGCGCAAAACGAGCTGGCGGA 30888
QY 454 LeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMet 473
|||...|||
Db 30889 ATTAGCGGCTCTTCATTAGAGGCTCAACAGTAGATTTATCCCGCAATTCGTTAACTTA 30948
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
|||...|||
Db 30949 ATGACATACCAAGTGGTTCCAGGCAATACAAAGTTATCCGTTGTCAGATGACGCTG 31008
QY 494 LeuGlnLysAlaLeuGluLeu 500
|||...|||
Db 31009 ATGAAACAAATGTGAACCTG 31029

RESULT 13
AB067197.06
Continuation (7 of 12) of AB067197 from base 600001 (Listeria innocua contig DNA sequenc
WP Sequence split into 12 fragments LOCUS AB067197 Accession AB067197
WP Fragment Name Begin End
```

```
WP AB067197_00 1 110000
WP AB067197_01 100001 210000
WP AB067197_02 200001 310000
WP AB067197_03 300001 410000
WP AB067197_04 400001 510000
WP AB067197_05 500001 610000
WP AB067197_06 600001 710000
WP AB067197_07 700001 810000
WP AB067197_08 800001 910000
WP AB067197_09 900001 1010000
WP AB067197_10 1000001 1110000
WP AB067197_11 1100001 1163020

Alignment Scores:
Score: 3e-27 Length: 110000
Pred. No.: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-i (1-502) x AB067197_06 (1-110000)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
|||...|||
Db 39538 ATGAATCAAACTATGATACAGCTATTCTGGGATGATGCGTTCACACAGCATATCA 39597
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
|||...|||
Db 39598 GTGACATCAAAATAATATGCAATTCGCAACACGACAGGATATAAAACAAAGCGTCGT 39657
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
|||...|||
Db 39658 TTCAATGATTTACTTTACCAAAACACAACTGGGATCTGTTCCAGCGGACTTTATGCTGGA 39717
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
|||...|||
Db 39718 ACMAAACCAATGAGTTCGGTTCGGTTCGAAATTTGGGCAATTTTAAACCGATTATACA 39777
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
|||...|||
Db 39778 GCAGGTTCTCCGACATCAACTGGCGAGAAACAAAGATGCAGACTGCAAGCGCGGCTTT 39837
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
|||...|||
Db 39838 TTCATTCTCGCATACGCTGGTGGGAATATCGTTTACACAGCAGACGCTAGCTTTGCA 39897
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
|||...|||
Db 39898 GTATCTGACAAATAACTATTTAACACGCAACGCAACGAAATACTGTTATGGGATATGCAACA 39957
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
|||...|||
Db 39958 GATAAAATGCGCAACGCTTTAAACGGAACCTTCAACCACTTCAATCCCATTAATAGC 40017
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
|||...|||
Db 40018 GCAATTCACGCGGAGCAACAAATAATGTTAGCGGTAAACATTCCTACTTGATGG 40077
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
|||...|||
Db 40078 GGGGAAAAA---GATACGATTTCTTCGAG----- 40104
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
|||...|||
Db 40105 -----CTT 40107
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
|||...|||
Db 40108 TCTGTATACGATATGCTGTGGGAAAAACATAAATTCAGTCAATATGAAAGCTGGACA 40167
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
|||...|||
```

Db 40169 CCAGATGCGAGTGGCAATGTTTCCTATGTAATGAATTAATCAATG-----GAC 40215
Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 40216 GGAAGCGTTAACTCTCA-----GTAACAGGAACACTTAATTAAC 40257
Qy 274 SerSerAsnGlyGluLeuLysMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 40258 AACGCGCAAGCGCAACTTACAAACCTGATGCA----- 40290
Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 40291 ---CITAAAAACATCAATCAATCCACTGTTACGGC----- 40326
Qy 314 PheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 40327 -----AAACAGTCAAT 40338
Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 40339 ATG-----GGTTAAACCTAAGTGGCTTAACA 40365
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 40366 AACTAC-----GGTACCAACCAAA 40383
Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 40384 GTATTCACCAACTCTCAGCGAAAGCGGCTGCACTGTAAGACATACCGACTTACC 40443
Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 40444 GATTCGTGTTATATTCAGTACGAGTACTCAGACGGTACAGTTATCCAGTTCGCCCACTT 40503
Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 40504 GCGGTGTCTACTTCTCCAAATGAAGACGCTTAGTCAAAATGGGAATGGCGAATATGTT 40563
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 40564 CAGGATATCTCTGAGATGAGTTCAGGCGTGTGTCGCAAAACGAGCTGCGGA 40623
Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 40624 ATTAGCGGCTCTTCAATAGAGCTCAACAGCTAGATTATTCGCGGAATTCGTTAACTTA 40683
Qy 474 IleIleIleGlaArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
Db 40684 ATGACATACCAAGTGTGTTCCAGCAATACAAAGTTATCCGTTGTCAGATGACGTTG 40743
Qy 494 LeuGlnLysAlaLeuGluLeu 500
Db 40744 ATGAACAAATGTGAACTTG 40764

RESULT 14
ABA03041_07
Continuation (8 of 30) of ABA03041 from base 700001 (Listeria monocytogenes EGD-e genome
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

Fragment Name	Begin	End
WP ABA03041_00	1	110000
WP ABA03041_01	100001	210000
WP ABA03041_02	200001	310000
WP ABA03041_03	300001	410000
WP ABA03041_04	400001	510000
WP ABA03041_05	500001	610000
WP ABA03041_06	600001	710000
WP ABA03041_07	700001	810000
WP ABA03041_08	800001	910000
WP ABA03041_09	900001	1010000
WP ABA03041_10	1000001	1110000
WP ABA03041_11	1100001	1210000
WP ABA03041_12	1200001	1310000
WP ABA03041_13	1300001	1410000

WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 2000001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528

Alignment Scores:

Pred. No.: 3e-27 Length: 110000
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x ABA03041_07 (1-110000)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 30516 ATGAATCAACTATGATATACAGCTATTTCTGGATGATGATCGTTCACCAAGCATTTATCA 30575
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 30576 GTAACATCAATAATATTTGCCAATGCCAACACACAGCATATAAAACAAAGCGTGGTT 30635
Qy 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 30636 TTCAATGATTACTTTTACCAAAACACAATGGGATCTGTCAGCGGACTTTATGCTGGA 30695
Qy 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 30696 ACAAAACCAATAGTTCGTTCCGGTTCGAAATTCGAAATTCGGCGGATTTTAACTGACTATA 30755
Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
Db 30756 GCAGGTTCGCCGACTTCACTGCGAGAAATAAAGATGACGACTACAAAGCCCGGGCTTT 30815
Qy 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 30816 TTCAATGCTGGTATACGCTGGTGGGAATATCGTTTATACGTCAGCGTAGCTTTGCA 30875
Qy 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 30876 GTATCTGACATAACTATTTTAAACAACGACGACGAGAAATACGTTATGGGTACGCAACG 30935
Qy 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 30936 GACAAAATGTAACGTTTAAACGGTAATTTACACCCATCCAAATTCGCTAATAATAGC 30995
Qy 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 30996 GCAATTCAGGGGAAGCAACAAAAATGCTAGCCTAAGCGGTACATCCCACTGATTGG 31055
Qy 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db 31056 GCGCAAAAA---GATACGATTCTTCGAG----- 31082
Qy 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 31083 -----CTG 31085
Qy 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233

PF 23-JUN-1998; 98WO-US013041.
 PR 24-JUN-1997; 97US-0050667P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Fraser CM;
 XX WPI; 1999-081273/07.
 PT New isolated Treponema pallidum nucleic acids - used to develop products
 PT for the detection, diagnosis, characterisation, prevention and therapy of
 PT T. pallidum infections, particularly syphilis.
 XX Claim 1; Page 491-497; 1150pp; English.
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection diagnosis,
 CC characterisation, prevention and therapy for T. pallidum infections,
 CC particularly syphilis. They can also be used for detecting diseases
 CC related to Borrelia infections in animals, and for the production of
 CC biosynthetic products such as enzymes
 XX Sequence 10461 BP; 2325 A; 3411 C; 2493 G; 2213 T; 0 U; 19 Other;

Alignment Scores:
 Pred. No.: 7,67e-28 Length: 10461
 Score: 476.00 Matches: 153
 Percent Similarity: 42.34% Conservative: 79
 Best Local Similarity: 27.92% Mismatches: 185
 Query Match: 18.41% Indels: 134
 DB: 2 Gaps: 14

US-10-009-823A-1 (1-502) x AAX20553 (1-10461)

QY 1 MetMetGlySerLeuPheLeuAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 2293 ATGATGGGTCACTTTTTCAGGTGTCTGATGAGATCATCAACGGCGATGAT 2234
 QY 21 ThrValSerAsnAsnLeuAlaAsnThrIleGlyTyrLysGlnGlnVal 40
 DB 2233 GTCATTGGGAACACGTGCGGAACGTTAACTACCTACCGGTTTAAAGCGTGGGTGTTAT 2174
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 DB 2173 TTTCAGATCTTATTCTCAGCACTGAGTGGCGGTGCGCGTCCGATCAAGAGTTGA 2114
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 DB 2113 GGAGTGAATCCCAAGGAAGTGGGATTTGGCGTCTGATTGCAAGCATCATGTTTCA 2054
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
 DB 2053 ACCGAGGTGCATGCAACGACGGGTATCAATACGATGTGTCTATTACGGGGAGTGT 1994
 QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 DB 1993 TTTTGTGTCTGAAAGTGGGGAAGACGTTTTCACCGGCGAGT-GCCTTTGGGGT 1935
 QY 116 ThrGlnAspGlyPheLeuAsnAspProSer-----Gly 126
 DB 1934 GATAATGCGGGCACTCTGTAACCTCGCAATGATGCGGTTCAGGTTGGATGGCG 1875
 QY 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlnThrLeuGlu 146
 DB 1874 CAGGACGTGGCGGGGAGCGTTTAAATTCCTCTGCACAGACGCGAGGATCTGTTATC 1815
 QY 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
 DB 1814 CCCATTGG-----GCAAAAGATAGATGCGCACAGACCGACG 1780
 QY 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186

DB 1779 ACTGTTCACTATGCTGTAAATTTTACAAAGCGTCTGCTGAGCTTGTCTGCAGATCGAAC 1720
 QY 187 ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206
 DB 1719 GAAGGGACGTGCGTAAGTCCACGTGG-----ACAACT 1687
 QY 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
 DB 1686 GACITTT-----CAAGTGTATGATAGTTCGGGACGACGATACGTTG 1645
 QY 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
 DB 1644 CAGATTAACTTTTCGCTGTCGCGGAGCG-----AACAACTAGTGGGACGACATGTC 1591
 QY 247 AlaMetAsnPro-----SerGluAsp 253
 DB 1590 GCAGTGGATCGGGGACAGAGGTAGATACGAAACGGGTGTAGGGGTGGGACATCTGAC 1531
 QY 254 GlySerAla-----AlaSerGlyThr 260
 DB 1530 GGTGCGCAACACCTTTATTGATAATTTTGGACACCTCGCTTCAGTGACT 1471
 QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 1470 GACACTGACGG----- 1459
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
 DB 1458 AACGTGACCGGT-----CCTACCGGACAGGTG----- 1432
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 DB 1431 -----CTCCTGAAGCGTGTACGATGTTGTCGTGGTGG- 1399
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 DB 1398 -----AATCCGACGATGACGAGGAGGTT 1375
 QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 1374 ACGCGCACGCTTTCACGCTCACTTGGGTGAAAT----- 1339
 QY 361 SerSerGlyAsnSerThrAlaArgAsn-----GlySerSerSerThr 374
 DB 1338 -----GGCACCGCGCGCAATACGATTACGCGATTGCTGAACGACGATCTACC 1291
 QY 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
 DB 1290 AAAGC-TACCGGACGAGGAGGTACGCGATGGGATATTGGAATAATTTAAATAGATCAA 1232
 QY 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
 DB 1231 AGCGTGTCTACTCTGTGTGTTTCAATGGGTGAGCCAGACATGCGCCAGCTCGCA 1172
 QY 415 LeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAla 434
 DB 1171 CTTGACGATTTCAAATCAAGGTGGTCTTTGAGAAGGACGAGAGACACCTACGTACAA 1112
 QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
 DB 1111 TCGAACAACTCAGGGATAGCAACATTAGCACCTCGGGGTGATGGGAGGGAAGTTG 1052
 QY 455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
 DB 1051 ATTGACGGACACTTGCAGATGACGACGATAGATTAAACCGATCAATTTACGGATATGATC 992
 QY 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
 DB 991 ATTACCCAAAGAGGTTTACGGCGGCGCAAGACGATTTCAGACATCAGACACCATGTTG 932
 QY 495 GlnLysAlaLeuGluLeuLysArg 502
 DB 931 GATACGCTGTTGAGTTTGAAGCGC 908

QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAen 333
Db 5556
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 5568 ATG-----GGCTTAAACCTAAGTGGCTTAACC 5594
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 5595 ARCTAC-----GGAAACAACCAA 5612
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 5613 GTATTCTCACCACCTCTGACGGTAAGCGCTGCGACTGTAAAGATTATGCACTTACC 5672
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 5673 GATTCTGGCTATATTCAGTGAGTACTCAGATGGTACAGTTATCCCACTTGCCTCAACT 5732
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 5733 CGCGTGGCTACTTCTCCATGAGACGGCTTACTCAAAATGGGACCGCGAATATGTT 5792
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 5793 CCAGGATTATCTCTGGCGATGCAGTATACGGGTTGCTGGCCAAATGGCGTGGCGGA 5852
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 5853 ATTAGCGGTTCTCTCTAGAAAGTTCAAACGTACACTTGTCCCGTGAATTCGTTAACTTA 5912
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 5913 ATGCATACCAAAATGTTTCCAAAGGCGCATACAAAGATTATTCGTGTGGCGACACGTTG 5972
QY 494 LeuGlnLysAlaLeuGluLeu 500
Db 5973 ATGAATCAAACTGTGCCTTG 5993

RESULT 19
ACF68483
ID ACF68483 standard; DNA; 1215 BP.
AC ACF68483;
XX 20-NOV-2003 (first entry)
XX Photobabidus luminescens nucleotide sequence #6950.
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX Photobabidus luminescens.
OS
XX
XX W0200294867-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.

XX Genomic sequence of Photobabidus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 6950; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photobabidus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX
SQ Sequence 1215 BP; 342 A; 257 C; 304 G; 312 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.58e-27 Length: 1215
Score: 448.50 Matches: 141
Percent Similarity: 40.75% Conservatives: 66
Best Local Similarity: 20.75% Mismatches: 174
Query Match: 17.34% Indels: 127
DB: 10 Gaps: 17

US-10-009-823A-1 (1-502) x ACF68483 (1-1215)

QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 22 AGTGGTTTGAATCGCGCGCAGGTAATTGGATCTTATTGGTAATAATATCGTAACGCG 81
QY 30 AsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 82 CGCATATATGTTTAAATCCAGCAGCGTGTCTTTCTGACATTTTGGC----- 132
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 133 -----GGTTCT-----CAGGTGGGACTGGGGTAAAGTATCG 165
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 166 GGTGTATATCAGAAATTTTAAAGATGGTACGCCAACGACGACTAACCGAACAATTGATCTG 225
QY 90 AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThr 108
Db 226 CGATCAGCGAGCGGGGTTTTTTCGTATGCAAGCGGTGATGCGGGATTTATATTC 285
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 286 CGTAACCGCCAATTCAAGATGATGAGAACCGCAATATATCAACATCGCAAGGATGAAA 345
QY 129 LeuMetGly-----SerArgIleSerAsnAsnProAsnIleLysLys---Glu 143
Db 346 TTGACCGGCTATCCCGCTGCCAGGTCATGTGGCGCGCGGAGATTCAGAAAGGGCT 405
QY 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer-----Pro 161
Db 406 AACCCAGAGCCCAAT-----ACAATACCTCAAGGATATATTGATG 444
QY 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181

		:::::	:::		:::::		----	TTC	CG	519
Db	445	GCTAAAGCAACACTAAGGTTTACGATGACAGCCAACCTT---	AATTGCATGCATAAAAA	501						
Qy	182	GLnSerGluAlaAsnProTyrrPheAlaLeuGluSerTrpLysGlyAsnGlyThrPro	201							
Db	502	CCGACAGAGCT-----								
Qy	202	ProileSerThrSerAsenTyrrSerTyrrAlaGlnProMetArgValTyrrAspGlnGlnGly	221							
Db	520	CCGGAAAATAACGATTCATTATAACTCACCAGTGTTACCACTTTGTAGCTGGGT	579							
Qy	222	AsnSerHisAspileThrValTyrrPheAspGlyAlaProSerSerThrGlySerLysThr	241							
Db	580	AACCCACATGCTATCCAGGTTTCTTT-----								
Qy	242	PheGluTyrrLeuValAlaMetAsnProSerGlu-----	AspGlySerAla	256						
Db	607	-----GTCAAACAATGATAAGCAATGGCAGGTTCATGCTCAGGATAGTCGC	654							
Qy	257	AlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsn	276							
Db	655	GTTCAGGCGCGCTAAAGCCGCTTCGTTG-----	GGGACGTTGAANAATTGATGGCAAC	705						
Qy	277	GlyGluLeu-----	LysAsnMetThrAlaPheThrProThrGlySerAlaThrLys	293						
Db	706	GGTAACATAATTGGTGACAATAATCAGTTTACATTTAAT-----	744							
Qy	294	AspLeuAsnAlaTrpGlnProAlaProLeuValAsnClyLeuProGlnPheSerAlaAsn	313							
Db	745	-----GTTCTCTCTCTGAAT-----	759							
Qy	314	PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn	333							
Db	760	-----GGTTCTTAATGCCGCTGATACACTATTGATTTAATGGCAGTAACAGCAAAAG	813							
Qy	334	MetTrpAlaGlyAlaProAlaSerAlaAlaAlleGlyThrAspIleGlyLysLeuPro	353							
Db	814	-----ATTGATATCAGACATATCAGCACCC	840							
Qy	354	SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer	373							
Db	840	-----	840							
Qy	374	ThrArgArgTyrrSerGlnAspGlyTyrrProGlnGlyAspLeuValAspValThrIleThr	393							
Db	841	-----AAGCAAGACGGTTATGCCGCGAGGTGAATTCAAGGGCTACCGCTATTGAG	888							
Qy	394	SerGluGlyLysLeuGlnGlyLysTyrrSerAsnSerGlnValValAspPheTyrrAsnIle	413							
Db	889	CAAGATGGCCGCTGTCATCGGCACCTATACCAATGACCAATCCCCAATTATTAGTCCAGATC	948							
Qy	414	ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrrSer	433							
Db	949	GTTCCTGACTAACCTTTGCTAAACCCAGAGGAGTGGCAGCAAGGGTGATAACGTTGGGTA	1008							
Qy	434	AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrrGlyLys	453							
Db	1009	GAGACGAGCCGCTCAGGTAACCCAGTGGTTGGCAATTCGGGATTCGGCGGTTTTGGCAAA	1068							
Qy	454	LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet	473							
Db	1069	TTGCTCAGTGGTTTCGCTGGAGGATCCAAAGCTGCATATGAGCCAGGAGCTGTTAATATG	1128							
Qy	474	IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet	493							
Db	1129	ATCGTGGCGGCGCAACTACCAATCCGTTGGCGCAACCAATTAACCCCAAGATCAGATC	1188							
Qy	494	LeuGlnLysAlaLeuGluLeuLys	501							
Db	1189	CTACAAACGCTGGTCAGCATGCGC	1212							

Qy 30 AsnThrIleGlyTyrIleGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 90512 GCGACATATGGTTTAAATCCAGCAGCGTGTCTTTCTGACATTTTGC 90462
Qy 50 AlaIleGlySerThrClySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 90461 -----GGTTCT-----CAGTGGGACTGGGGTAAAGTATCG 90429
Qy 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 90428 GGTGTTAATCAGAAATTTAAAGATGTATCGCCCAACGACGACTAACCGAACACTTCATCTG 90369
Qy 90 AlaIleGlyGlySerPheGlnValThrLeuGluAsp---LysValHisThr 108
Db 90368 GCGATCACCGAGCGGGTGTTCGTATGCAAGCGGTGATGCGGGATTTATTTATTC 90309
Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 90308 CGTAACGGCCAATTCAGATGATGATGAGAACCGCAATATTATCAATGATCAAGGCATGAAA 90249
Qy 129 LeuMetGly-----SerArgIleSerAsnAsnProAsnIleLysLys---Glu 143
Db 90248 TTGACCGGTATCCCGCTCCAGCGTCAATGTGTGGCGCGCGAGATTTCAGAAAGGGCT 90189
Qy 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer-----Pro 161
Db 90188 AACCCAGAGCCAAAT-----ACATACCTCAAGTATATGTATG 90150
Qy 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
Db 90149 GCTAAAGCAACAACTAAGGTACGATGACAGCAACCTT---AATTCGATGATAAAAA 90093
Qy 182 GlnSerGluAlaAsnProThrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
Db 90092 CGACAGAGGCT-----TTCTCG 90075
Qy 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGly 221
Db 90074 CCGGAAATAACGATCTTATACTATCTACCAGTGTACCAGCTTTCATAGCTGGT 90015
Qy 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241
Db 90014 AACCCACATGCTATCCAGTGTCTCTTT----- 89988
Qy 242 PheGluTyrLeuValAlaMetAsnProSerGlu-----AspGlySerAla 256
Db 89987 -----GTGAAACGATGATAACGAATGCGAGTTCATGCTCAGGATAGTGG 89940
Qy 257 AlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsn 276
Db 89939 GTTGAGGCGCTAAAGCGCTTCGTG-----GGGACGTTGAATTTGATGCGCAC 89889
Qy 277 GlyGluLeu-----LysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 89888 GGTAACTAATGCGTGAATATACGTTACATTTAAT----- 89850
Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 89849 -----GTTCTCTCTCTGAAT----- 89835
Qy 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsn 333
Db 89834 -----GGTTCTAATGCGGTGATACATCATTTGATTTAATGGCAGTAAACAGCAAAAG 89781
Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 89780 -----ATTGATACTGACGATATCAGCACCC 89754
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 89754 ----- 89754

Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 89753 -----AAGCAAGACGGTTATGCGCGAGGTGAATTCAGGCGCTACCGTATTGAG 89706
Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 89705 CAAGATGCGCGTGCATCGGCACCTATACCAATGAGCAATCCCAATATTATTAGGTACATC 89646
Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 89645 GTTCTGACTAATTTGTAACCCAGAGGACTGCGACGAGGAGGTGATAACGTTTGGGTA 89586
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 89585 GAGACGAGCGCTTCAGGTAACCCAGTGTGGCATTCGCGATTCTGGCGGTTTGGCAA 89526
Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 89525 TTGCTCAGTGGTTCGCTGGAAGCATCCACGTCGATATGAGCCAGGAGCTGGTTAATATG 89466
Qy 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 89465 ATCGTGGCGAGCGCAACTACCATCCAGTGGCGCAACCATTAACCCAGATCAGATC 89406
Qy 494 LeuGlnLysAlaLeuGluLeuLys 501
Db 89405 CTACAAACGCTGGTCAGCATGCGC 89382

RESULT 21

ACF67367_12/c Continuation (13 of 57) of ACF67367 from base 1200001 (Photorhabdus luminescens nucleotic
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

Fragment Name	Begin	End
WP ACF67367_00	1	110000
WP ACF67367_01	100001	210000
WP ACF67367_02	200001	310000
WP ACF67367_03	300001	410000
WP ACF67367_04	400001	510000
WP ACF67367_05	500001	610000
WP ACF67367_06	600001	710000
WP ACF67367_07	700001	810000
WP ACF67367_08	800001	910000
WP ACF67367_09	900001	1010000
WP ACF67367_10	1000001	1110000
WP ACF67367_11	1100001	1210000
WP ACF67367_12	1200001	1310000
WP ACF67367_13	1300001	1410000
WP ACF67367_14	1400001	1510000
WP ACF67367_15	1500001	1610000
WP ACF67367_16	1600001	1710000
WP ACF67367_17	1700001	1810000
WP ACF67367_18	1800001	1910000
WP ACF67367_19	1900001	2010000
WP ACF67367_20	2000001	2110000
WP ACF67367_21	2100001	2210000
WP ACF67367_22	2200001	2310000
WP ACF67367_23	2300001	2410000
WP ACF67367_24	2400001	2510000
WP ACF67367_25	2500001	2610000
WP ACF67367_26	2600001	2710000
WP ACF67367_27	2700001	2810000
WP ACF67367_28	2800001	2910000
WP ACF67367_29	2900001	3010000
WP ACF67367_30	3000001	3110000
WP ACF67367_31	3100001	3210000
WP ACF67367_32	3200001	3310000
WP ACF67367_33	3300001	3410000
WP ACF67367_34	3400001	3510000
WP ACF67367_35	3500001	3610000
WP ACF67367_36	3600001	3710000
WP ACF67367_37	3700001	3810000
WP ACF67367_38	3800001	3910000
WP ACF67367_39	3900001	4010000

AC67367_40 4000001 4110000
 AC67367_41 4100001 4210000
 AC67367_42 4200001 4310000
 AC67367_43 4300001 4410000
 AC67367_44 4400001 4510000
 AC67367_45 4500001 4610000
 AC67367_46 4600001 4710000
 AC67367_47 4700001 4810000
 AC67367_48 4800001 4910000
 AC67367_49 4900001 5010000
 AC67367_50 5000001 5110000
 AC67367_51 5100001 5210000
 AC67367_52 5200001 5310000
 AC67367_53 5300001 5410000
 AC67367_54 5400001 5510000
 AC67367_55 5500001 5610000
 AC67367_56 5600001 5648894

Alignment Scores:
 Pred. No.: 2,418-24 Length: 110000
 Score: 448.50 Matches: 141
 Percent Similarity: 40.75% Conservative: 66
 Best Local Similarity: 27.76% Mismatches: 174
 Query Match: 17.34% Indels: 127
 DB: 10 Gaps: 17

US-10-009-823A-1 (1-502) x ACF67367_12 (1-110000)

QY	10	ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
DB	27341	ACTGGTTTCAGTCCGCGCAGGTAATTGGATGTTATGGTAATAATATCGTAACGCG 27282
QY	30	AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
DB	27281	GGGACATATGGTTTAAATCCAGCAGCGTCTTTTCTGACATTTTGGCC----- 27231
QY	50	AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
DB	27230	-----GGTTCT-----CAGTGGGACTGGGGTAAAGTATCG 27198
QY	70	SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
DB	27197	GGTGTAAATACGAATTTTAAAGATGGTACGCCAACGACACTAACCGAACACTTGATCTG 27138
QY	90	AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAsp--LysValHisTyrThr 108
DB	27137	GGGATCACCGCAGCGGTTTTCGGTATCGAAGCGCGTGTATCGGGGATTATATTC 27078
QY	109	ArgAlaGlyAsnArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
DB	27077	CGTAACGCCCAATTCAGATGGATCAGAACCGCAATATTATCAATGCAAGGCATGAA 27018
QY	129	LeuMetGly-----SerArgIleSerAsnAsnProAsnIleLysLys--Glu 143
DB	27017	TTGACCGGCTATCCGTCGCCAGCTCAATGGTGGCGCGCGGAGATTTCAGAAAGGGCT 26958
QY	144	ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer-----Pro 161
DB	26957	AAACCAGAGCCAAAT-----ACAATACCTCAAGGTATATTGATG 26919
QY	162	AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
DB	26918	GCTAAGCAACAACTAAGTTACGATGACGCCAACCTT---AATTCGATGATATAAANA 26862
QY	182	GlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyrLysGlyAsnGlyThrPro 201
DB	26861	CCGACAGAGGCT-----TTCTCG 26844
QY	202	ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGly 221
DB	26843	CCGGAANAATAACGATCTTATTAATATCTTACCACGATGTTACCACTTATGCTGGGT 26784
QY	222	AsnSerHisAspIleThrValTyrPheAspGlyValaProSerSerThrGlySerLysThr 241

Pred. No.:	1.33e-21	Length:	110000
Score:	414.50	Matches:	132
Percent Similarity:	39.96%	Conservative:	73
Best Local Similarity:	25.73%	Mismatches:	187
Query Match:	16.03%	Indels:	121
DB:	6	Gaps:	13

US-10-009-823A-1 (1-502) x ABA92787_3 (1-110000)

QY	4	SerLeupheileGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer	23
DB	72153	TCAAATATGATGATCTAAGTGGCTTACTAGCAAAATAATGATTACATGGAATTTATATCC	72212
QY	24	AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp	43
DB	72213	ANTAATATTGCTAACGCATCACTAGGATATAATCTCGTAAACCTCTTTTGTAT	72272
QY	44	LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly	63
DB	72273	ATGTTTCTCATCTATTTATTCAATACTACTAAT	72311
QY	64	MetGlyAlaGlnValGlySerValAsgThrIlePheThrGlnGlyAlaPheGluProGly	83
DB	72312	TACGGAGTGGGTATTTCAAGTATATACAAACTTTTAATAATGGCATGTAGTTGAACT	72371
QY	84	AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGln---ValThrLeuGlu	102
DB	72372	GGACGAGATTGGATTAGGAATTATAAAGACGGCTTTTTCGCTCTTGACACAGTCAA	72431
QY	103	AspLysValHisIleThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn	122
DB	72432	GGTCATGCTTATATACAGAGATGGGCAATCTTCTCGATTAAGATCAAAATATATC	72491
QY	123	AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn	136
DB	72492	AATATTCAAGGTATGTATCTAACTGGACTTAATACATCTTGTTCAAAAAGTGATTTAAT	72551
QY	137	AsnProAsnIleLysGlyThrLeuGluProIleGlnLeuAspPheAsnAspProThr	156
DB	72552	AAT-----AGATCCAACTTAGAACCTTATTAATTTA-----AAAAATCTAAT	72593
QY	157	ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly---	175
DB	72594	ATTTTAAAAACAACCTACTCTGAAATCATGCTTAAAGCGTTTTTGATCGTAACTATC	72653
QY	176	-----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe	189
DB	72654	GAATCAAAAAGCAGTGTGGATAATCTGCACAACTATCTAAACAGAGATTACATG	72713
QY	190	AlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSer	209
DB	72714	ACTTATATTAGC-----	72725
QY	210	TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr	229
DB	72726	-----ATATATAATAAGAGGGAAGAAAAAGAGATATTACTGTTTCT	72767
QY	230	PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn	249
DB	72768	TTTAAT-----AAAAAGGAACAATAAATGACAGATTAAT-----GTGGAATCAAT	72815
QY	250	ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly	269
DB	72816	GATTCGATGATAAGAGACTATAAAAAAT-----AGTTTC	72851
QY	270	ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly	289
DB	72852	GATTTACGTTTAAATGATGATGCGAATTAACTCTGATAATGTTTTTAATATTACATCT	72911
QY	290	SerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGln	309
DB	72912	AAAGATTCTAAAAAG-----	72926

310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLys 329
72927 -----TATGAAATATCATCTTTTAAATTTA----- 72950
330 SerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349
72950 ----- 72950
350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
72951 -----ACAGGTACTATAGAACATCAAT 72974
370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
72975 TCTGATGTTCTTCTGGGAAGAACATCTCAAAACGATACCTCAAGGTAAATTTAAAAACA 73034
390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
73035 TTTGATATTGTTACTAATGGTGAATATTATGGAACATATTGCAATCAAAACACAAACA 73094
410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsn 429
73095 ATAGTCAAAATATTATTATCAAAATTTATCAATCCAGAAAAATTTACAACCTGAAAGTGGT 73154
430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu---ProGlyThr 448
73155 AATTATGGTCTGCTACTGCAAAATCAGGTGAGCAAAACAGCAATGAAAGCGGTATT 73214
449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
73215 CAAGAATCAGGAGTGTAAAGCAATAAAACGCTAGCAAGTATCAATGTTGATTGCAATAAA 73274
469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
73275 GAATTAATCAATATGATTATAGCAACAGTAAATTTATCAATCTAACGCTCAATCTTTTAAA 73334
489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
73335 ACAGAGATAAATAAATAATTAATACATTAAATTTACAG 73373

RESULT 23
AAS88896
ID AAS88896 standard; cDNA; 1912 BP.
XX AAS88896;
AC AAS88896;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24700.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABC24709.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

[illegible]

QY 369 AsnGlySer-----SerSerThrArgArgTyrSerGlnAsp 380
 DB 94714 GAAGAGATGATTTTACAAATTTCTTACCTGATGCACTTATGATATATAGAT 94773
 QY 381 Gly----- 381
 DB 94774 GGGTCATTAAATCGATTCTTAATCGAGAGCTTGTAAACCAAGCATACAAGTATTG 94833
 QY 382 -----TyrProGlnGlyAspLeuValAsp---ValThrIleThrSerGluGly 396
 DB 94834 CCTAATATATCTTCCAGAGAAATATATCCAAATCAATATCTGAAGAGGA 94893
 QY 397 LysLeuGlnGlyLysTyrSerAsnSer---GlnValValAspPheTyrAsnIleProLeu 415
 DB 94894 ATAGTATCGGTAAATATGATACCAAGCAACCAATAGAGCTTGGCAATTGAATA 94953
 QY 416 AlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThr 435
 DB 94954 TCAAGATTATCAATCTCGAGGACTAAGTCCATTGGAAGCAATTTATTTAAAGAAACA 95013
 QY 436 LeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer 455
 DB 95014 GCTGGATCAGGCCCAAGAAATAGCAGGAATACCAAGAAAGTGAAGCACTAAGG 95073
 QY 456 ValAsnGlnGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle 475
 DB 95074 CAAGGCATCTTGAATGTCAATGTATCTATTGCTGAAGAAATGGTAACATGATAGTA 95133
 QY 476 IleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGln 495
 DB 95134 GCTCAAGGGCTTATGAATAAATCAAAAGCTATTCAAACTTCTGACATATGTTAGGA 95193
 QY 496 LysAlaLeuGluLeuLysArg 502
 DB 95194 ATTGCAAACTAATTTAAAGG 95214
 RESULT 25
 ID ABD04540 standard; DNA; 810 BP.
 XX
 AC ABD04540;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Pseudomonas aeruginosa polynucleotide #3144.
 XX
 KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
 XX
 KW antibacterial.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN US5551795-B1.
 XX
 PD 22-APR-2003.
 XX
 PF 18-FEB-1999; 99US-00252991.
 XX
 PR 18-FEB-1998; 98US-0074788P.
 XX
 PR 27-JUL-1998; 98US-0094190P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
 XX
 DR WPI: 2003-615309/58.
 XX
 DR P-PSDB; ABO70969.
 XX
 PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 XX
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 XX
 PT pathological conditions resulting from bacterial infection.
 XX
 PS Disclosure; SEQ ID NO 3144; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 810 BP; 166 A; 299 C; 226 G; 119 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.35e-15 Length: 810
 Score: 307.00 Matches: 108
 Percent Similarity: 30.74% Conservative: 46
 Best Local Similarity: 21.56% Mismatches: 93
 Query Match: 11.87% Indels: 254
 DB: 11 Gaps: 11
 US-10-009-823A-1 (1-502) x ABD04540 (1-810)
 QY 1 MetMetGlySerLeuPheIleGlyAlaThrHisSerThrGlyLeuGly 20
 DB 25 ATGTATCGGCACGTGGTCTGCGCCAGACCGTCTGTCGCCAGCATGAACCTGACC 84
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 85 ACCATTTCCAAACACCTGGCCCAAGCTATCCACACCGGCTTCAAGCGCGCGGAG 144
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 145 TTCAGGACCTGTGTACCATCGCGCCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 204
 QY 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79
 DB 205 GAGCTGCTTCGGGCGTCAACTGGTACCGGTGTGGCGCTGTGGCGCACCCAGAGATC 264
 QY 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94
 DB 265 TTCACCCGGGCGAGCGTCGACAGCACCGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 324
 QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 DB 325 GGCCTCTTCAGGTCCTGCTCGCGAGCGGACCGGTGCTTACACCCGCGAGCGGCGGCTC 384
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 DB 385 CACCTGAACCTCCGAGCGGCGAGATCGTCACTCCACCGGCTTCGCCCTG----- 432
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 432----- 432
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 DB 433 GAGCCAGCGGATC----- 444
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
 DB 444----- 444
 QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 DB 444----- 444

QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 444 -----
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 445 -----GTGGTGCCCAACGAG--- 459
QY 254 GlySerAlaAspGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 460 -----ACCCAGACCTTC 471
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 472 ACC----- 474
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 474 ----- 474
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 475 ---GTCGCG----- 480
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 480 ----- 480
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 480 ----- 480
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 481 ---CAGGACGCG-----ACCGTCTCGGTGAC 504
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 505 ACCACCGCGCAACGCCAG-----CCGCGGTGATCGCG-----AACATC 543
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 544 CAGACCGCGCACTTCATCAACCGCGCGCTCGAGCGCATCGCGCACTCGCAACCTGTTCTCG 603
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 604 GAAACCGCGCTCCAGCGCGCGCGCGAGTGGTACGCGCGGTCTCAACGCGCTCGCGACG 663
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 664 GTTGCCCGAGAACACCCCTGGGAAACTCCACGTCACGTCGAGGAACTGGTGAACATG 723
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
Db 724 ATCACCACCGCGCGCTACGAGATGAATCCAGTCACTCCACCGCGCGACAGATG 783
QY 494 Leu 494
Db 784 TTG 786
RESULT 26
ABD04219/c
ID ABD04219 standard; DNA; 1545 BP.
XX AC ABD04219;
XX AC
XX DT 29-JUL-2004 (first entry)
XX DE Pseudomonas aeruginosa polynucleotide #2823.
XX KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX KW antibacterial.

XX OS Pseudomonas aeruginosa.
XX US6551795-B1.
XX 22-APR-2003.
XX 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX P-PSDB; ABO70648.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, of
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX Disclosure; SEQ ID NO 2823; 455pp; English.
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide, of
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using biochip technology. Sequences ABD01397-
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX SQ Sequence 1545 BP; 226 A; 447 C; 568 G; 304 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 7,66e-15 Length: 1545
Score: 302.00 Matches: 107
Percent Similarity: 30.60% Conservative: 46
Best Local Similarity: 21.40% Mismatches: 93
Query Match: 11.68% Indels: 254
DB: 11 Gaps: 11
US-10-009-823A-1 (1-502) x ABD04219 (1-1545)
QY 2 MetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThr 21
Db 1545 TTATCGCACTGTGGTGCAGCAAGCCGCTGTCCGCCAGGACATGACCTGACAC 1486
QY 22 ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe 41
Db 1485 ATTTCCAAACACCTGGCCCAACGATATCCACCACCGCTTCAAGCGCGCGGAGTTC 1426
QY 42 GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln 61
Db 1425 CAGGACCTGTGTACCATGATCCGGCGCGCGCGCGCGCGCGCGAGCAGGAG 1366
QY 62 AlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAlaPhe 80
Db 1365 CTGCCTTCGGGCGCTCAACTGGGTACCGGTGCGGTGCGGTGCGGCGCGGATCTTC 1306
QY 81 GluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLysGly 95
Db 1305 ACCCG 1246

96 PhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPheArg 114
 1245 TTTCTCCAGGTCCTGCTCCGCGACGCGACCGTGTCTACACCGCGACGCGACGTTCCAC 1186
 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
 1185 CTGAACCTCCGACGGCAGCATGTCACCTCCACCGCTTCGCCCTG----- 1141
 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 1140 -----GAG 1138
 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
 1137 CCAGCGATC----- 1129
 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
 1129 ----- 1129
 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
 1129 ----- 1129
 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
 1129 ----- 1129
 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
 1128 -----GTGGTGCCCAACGAG----- 1114
 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
 1113 -----ACCGACCTTCACC 1099
 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
 1099 ----- 1099
 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
 1099 ----- 1099
 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsnMet 334
 1098 GTCCGC----- 1093
 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
 1093 ----- 1093
 355 MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr 374
 1093 ----- 1093
 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
 1092 -----ACCGTCTCGGTGACCAACC 1066
 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
 1065 ACCGCAACGCCACG-----CCGAGGTGATCGCG-----AACATCCAG 1027
 415 LeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAla 434
 1026 ACCGCGGACTTCATCAACCGCGCGCTGCGAGGCTATCGGCAACAACTGTTCTCTCGAA 967
 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
 966 ACCGCTCAGCGCGCGCGAGTCTGCGGTCTCAACGGCTCGGACCGGTT 907

455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
 906 GCCCAGAACACCTCGAATACTCCACCTCAACGTGTCGAGGAACTGTGAACATGATC 847
 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
 846 ACCACCCAGCGCGCTACGAGATGAATCCAAAGTTCATCTCCACCGCGACGATGTTG 787

RESULT 27
 ACF68481
 ID ACF68481 standard; DNA; 783 BP.
 XX
 AC ACF68481;
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens nucleotide sequence #6948.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough; gene; ds.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taouit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 XX
 WIPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 FS Claim 2; SEQ ID NO 6948; 1205pp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens genes
 XX
 SQ Sequence 783 BP; 235 A; 163 C; 178 G; 207 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1-2e-14 Length: 783
 Score: 295.00 Matches: 104

Percent Similarity: 30.53%
Best Local Similarity: 20.35%
Query Match: 11.41%
DB: 10
Conservative: 52
Mismatch: 93
Indels: 262
Gaps: 8

US-10-009-823A-1 (1-502) x ACF68481 (1-783)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 1 ATGATCCGATCATATGATGCTTAAACCTGGGCTGGATCCAGCAAAACCAATATGGAT 60
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 61 GTGATCCCAACAACTTGGCAACGCTGCACCAACATGCTTTTAAACCCAGCGTGCAT 120
QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGlySer 56
DB 121 TTTGAGATCTGCTTTACCAAGACATTCGTCAACAGGGCGGATGTCATCGGACAACT 180
QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
DB 181 CGCTTACCTTCTGTTTACAGATAGTACTGGGTACGCGGTAGCAACAGACGTATA 240
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
DB 241 CATAGTCAGGAACTTATCTGAAACTAATAACAGCCCAATAATGTTGCCATTAGAGGAAAG 300
QY 95 GlyPhePheGlnValThrLeuGluAspLysValHis---TyrThrArgAlaGlyAsnPhe 113
DB 301 GGTTCCTTCAGGTCTAGTCTACCTGATGGTACCAAGTCTTATACCGGTGACCGTTCATT 360
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
DB 361 CAGGAAGATCAAAACGACAGTGTGACCAACCGCTAACGGTTTTTGTATCTTC----- 411
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
DB 411 ----- 411
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValAsn 173
DB 412 ---CCTACTATT----- 420
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
DB 420 ----- 420
QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
DB 420 ----- 420
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
DB 420 ----- 420
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
DB 420 ----- 420
QY 254 GlySerAlaLaserGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 420 ----- 420
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 420 ----- 420
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 420 ----- 420
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333

DB 420 ----- 420
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 420 ----- 420
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
DB 421 -----ACCATTCCTGATAATGCGAGC----- 441
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 442 -----GACCTCACCATTAGC 456
QY 394 SerGluGly-----LysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
DB 457 AGTGATGTTATGTGTCAGTGTGAAGATGACAGGGCAATCGGTACCTCAGCAAGTTGGTCAA 516
QY 410 PheTyrAsnIleProLeuAlaArgPheThrSerGlnAspGlyLeuArgArgGluGlyAsn 429
DB 517 TTT-----ACTCTGACAACTTTTATCAATGACAGTGGTTAGAAAGTATCGGTCAA 567
QY 430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSer 449
DB 568 AACCTTTTATGTAGAGACTAGCAGCTCTGCGCTACCGATTGAAATATACACCGGGAATCAAT 627
QY 450 AsnTyrGlyLysLeuSerValAsnGlnLeuThrSerAsnValAspMetSerArgGlu 469
DB 628 GCGCAGGATGCTCCACGATAAATATGTGAAACATCTAACGTCATATCGCGGAAGA 687
QY 470 MetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThr 489
DB 688 TTGGTCAATATGATTGATTCAGACTCAGCGTCTTATGAAGTTAATAGTAAAGCGGTGCGACG 747
QY 490 AlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
DB 748 TCTGATCAGATGCTACNAAACTGACTCAGTTA 780

RESULT 28

AAK20515/c
ID AAK20515 standard; DNA; 32768 BP.

XX AAK20515;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of Treponema pallidum.

XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.

XX Treponema pallidum.

XX WO9859034-A2.

XX 30-DEC-1998.

XX 23-JUN-1998; 98WO-US013041.

XX 24-JUN-1997; 97US-0050667P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fraser CM;

XX WPI; 1999-081273/07.

XX New isolated Treponema pallidum nucleic acids - used to develop products
PT for the detection, diagnosis, characterisation, prevention and therapy of
PT T. pallidum infections, particularly syphilis.

XX Claim 1; Page 261-279; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection, diagnosis,
 CC characterisation, prevention and therapy for T. pallidum infections,
 CC particularly syphilis. They can also be used for detecting diseases
 CC related to Borrelia infections in animals, and for the production of
 CC biosynthetic products such as enzymes

XX
 SQ Sequence 32768 BP; 8253 A; 9783 C; 7257 G; 7433 T; 0 U; 42 Other;

Alignment Scores:

Pred. No.: 2,7e-12 Length: 32768
 Score: 291.00 Matches: 145
 Percent Similarity: 34.37% Conservativity: 77
 Best Local Similarity: 22.45% Mismatches: 196
 Query Match: 11.25% Indels: 229
 DB: 2 Gaps: 22

US-10-009-823A-1 (1-502) x AAX20515 (1-32768)

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 30751 ATGATCCGAGGGTGTATATCCGAGCCAGTGGCATGAGCGGCAGCAGCGCTGGAT 30692
 QY 21 ThrValSerAsnAsnLeuAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 30691 GCTATTGGCGAAACCTAGCGAATGTAGACACAAACAGCTACAGCGAGATGTAGCAGTT 30632
 QY 41 -----PheGlnAspLeuPheSerGlnAspLeu----- 49
 DB 30631 CACAGAGTTTTCAGAGCTCTATTGGCTGCTGTAAACGACGATGTGTGTCAGAAC 30572
 QY 50 AlaIleGlySerThrGly-----SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
 DB 30571 CCTTTCGGTCTCGGAGTTCTCTATAGTTGGAAGTTAGGCTCGGTGTAGGGTG 30512
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 DB 30511 AATGAACTGTTCACTGAATTTGACAGGGTTCGCTTAAACAAACGCAAACTCCTTCGGAT 30452
 QY 89 LeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
 DB 30451 ATTGCGCTTGAGGGATGGGTTTTTTTGTGATTAGGACTCCACAGGAGAGAGTACACC 30392
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 30391 CGGAATGGGAATTTCTAGTCGGTGTAGAGGTTACCTTATGACTAAAAATGGGTATCCA 30332
 QY 129 LeuMetGlySerArg----- 137
 DB 30331 GTACTGGAGAAATGGTCCACTTTTCTGAGGAGGATACATACATCAATCAAAAC 30272
 QY 138 ProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp-----PheAsnAsp--- 154
 DB 30271 GGGGAGATTATGTACGTCCCATAGATAGGCGACAGCTTGACGGTTTTTTTTTTAGATCGG 30212
 QY 154 ----- 154
 DB 30211 CTGAAATTTGACGTTTGAATAATGTTGCTTACCTGCAAAAGAGGGTGTGATACGTAT 30152
 QY 155 -----ProThrValAlaLysSerPro----- 161
 DB 30151 ATGCAAACTCCTGTTTCAGGGGCCCTTATTGCTCGGAAGTCTCTGAACGTCGGGTGCA 30092
 QY 162 -----AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly 175
 DB 30091 GTGCAGGGTTTTGTGAGCGCTGAATGTAACGTGTCATGAAATGGTTCTGATGATT 30032
 QY 176 AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla-LeuLeuGluSerTr 195
 DB 30031 GAGGTGAATCGCGGTATGAGGCGAACCCAAAAAATATTCAAGCAGAGAGATGGAATGATG 29972
 QY 195 PysGlyAsnGlyThr----- 200

DB 29971 GGCAGATTGTGGAACGAAAGTGGTGGCTGCAAAAGTAGTTGTCCGCATGTTTTTATAGAATGT 29912
 QY 201 -----ProProIleSerThrSerAsn-----TyrSe 209
 DB 29911 AAAGAAGATATAGACGACATAGCGTGTAGCGGAGTAGAGGGAGGTGTGAATGTGTACGA 29852
 QY 209 rTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTy 229
 DB 29851 ATTTGTGGAC-----CGTGTGCTA 29834
 QY 229 rPheAspGlyAlaProSer-----SerThrGlySerLysThrPheG1 243
 DB 29833 CGGGGATGAACCGCCAGCAGCAACATCGATACGTTAGCGGAATAATCTACCGAATGTGA 29774
 QY 243 uTyrLeu-----ValAlaMetAsnProSe 251
 DB 29773 ATACGTCTGGTTTAAAGACAGCGCTCGGAGTTCGAGGACCTGTTGTATCAGACCATA 29714
 QY 251 rGluAspGlySerAlaAlaSer-----GlyThrAspSerAl 263
 DB 29713 GCACCGCGGTAGCCTGCAACTCAGGACACACATTACCCCTGTTGGGTACCGGATGGGC 29654
 QY 263 a-GlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn--- 281
 DB 29653 ACGGGTGAACCTCCTCAACGACGCKTCTTCGACACAGGTTTCGTCACGACACCG 29594
 QY 282 -----MetThrAlaPheThrP 287
 DB 29593 GTGTGAGCGGAGATGCTGCTATTGCGGTGAGGGTTTTTTCGCGTACTGCAATATGACG 29534
 QY 287 roThrGlySerAlaThrLysasp-----LeuAsnAlaTrpGlnProAlaProL 303
 DB 29533 GAACCTATGCTTATACCCGCGAGCGTTCGTTCAAGTGGATGCGGAGCGCAATTGGTGA 29474
 QY 303 euValAsnGlyLeu-----ProGln-----PheSerAlaAsnPheValGlyAlaG 318
 DB 29473 CCTCAACCGGTTGCTGCTCTGTTCCAGAGATTATTTTCCGAAACTACGTAGAACACA 29414
 QY 318 lYleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyA 338
 DB 29413 GTATC----- 29409
 QY 338 laProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProI 358
 DB 29408 -----GCTATTAGCAAGAT----- 29394
 QY 358 leGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrS 378
 DB 29393 -----GGGCGGTAGCGTTCGAGTCCGGGAA----- 29367
 QY 378 erGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysL 398
 DB 29366 -----CAGGAGATCCGTTGAGTA-----GGACAGC 29339
 QY 398 euGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgp 418
 DB 29338 TTGAG-----CTGTA-TNCT 29325
 QY 418 heThrSerGluAspGlyLeuArgGlyGlyAsnAsnHisTyrSerAlaThrLeuAsps 438
 DB 29324 TCGCAAACTAGCGGGTTTCGACGAGAGGGGGGAAATGTTTACCAACACACTGGCT 29265
 QY 438 erGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnG 458
 DB 29264 CCGGCCAGCAATTCGCGAAGCGGGTTCGAGGATTCGAAAGTTCGAAAGTTCGAGCAAGT 29205
 QY 458 lnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnA 478
 DB 29204 TTTTGGAAATGTCAAACTGTCTCAGTAAGTGAATGGTGAACATGATTGTCCCCAGC 29145
 QY 478 rGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAla 498
 DB 29144 GTGCGTATGAGTTTAATTCAAAGCAATTCAAACGAGTGACAACTGCTCGGCACAGCAG 29085

498 euGluLeuLysArg 502
 29084 TGGCGCTCAAGCGT 29071
 RESULT 29
 AAT67783
 ID AAT67783 standard; DNA; 816 BP.
 AC AAT67783;
 XX
 XX 29-JUL-1997 (first entry)
 DT
 DE H. pylori flagella associated protein ORF 29298130.aa.
 DE
 KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; flagella;
 KW associated; flagellum; ds.
 XX
 XX Helicobacter pylori.
 OS
 XX Key Location/Qualifiers
 FH 1-816
 FT CDS /*tag= a
 FT
 XX WO9640893-AL.
 XX
 XX 19-DEC-1996. 96WO-US009122.
 XX
 XX 06-JUN-1996; 96WO-US009122.
 XX
 XX 07-JUN-1995; 95US-00487032.
 PR
 XX 01-APR-1996; 96US-00630405.
 XX
 XX (ASTR) ASTRA AB.
 PA
 XX Smith D, Berglindh OT, Mellgaard BL;
 PI
 XX WPI; 1997-052306/05.
 DR
 XX P-PSDB; AAW20373.
 DR
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 PT
 XX Claim 1; Page; 1481pp; English.
 PS
 XX The present sequence encodes a Helicobacter pylori flagella associated
 CC protein. The protein may be used in a vaccine to prevent or treat H.
 CC pylori infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial DNA.
 CC The sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts. Note: This DNA sequence is not
 CC reproduced in the specification and has been derived from the related
 CC specification, WO9719098
 CC
 XX Sequence 816 BP; 237 A; 199 C; 183 G; 197 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.65e-14 Length: 816
 Score: 288.00 Matches: 107
 Percent Similarity: 28.68% Conservatives: 39
 Best Local Similarity: 21.02% Mismatches: 109
 Query Match: 11.14% Indels: 254

US-10-009-823A-1 (1-502) x AAT67783 (1-816)
 QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 28 ATGCTCCGCTCTCTATAGTGCACCTTCAGGATGCTCGCCCAACAAACGCAATTCAC 87
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 88 ACCACTTCAAAACAACATCGCCATGTCAATACCAACCGGGTTTAAAAAATCTCGCGCAT 147
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 148 TTTAACGACTTGTTTTACCAAGCATGCATATACCCCGCACCAACACAAAGCAACACGACT 207
 QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 DB 208 TTATGCCAGATGCGATCGAAGTGGCTTGGCTAGCGCTAGTCGATTAACAAAATG 267
 QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 DB 268 TTTTCGCAAGCGACGCCCTTAAAGAAACGGAGATAATTTAGATATTGTATTACAGTAAA 327
 QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 DB 328 GCGTTTTTCAAGTCCAGCTTCTGTATGGCACTACCGCTTACACAGGAGCGGGAATTC 387
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 DB 388 AAGCTAGACGACGACGAGGCAATCTTGTAAACAGCGAGGCTATCTCTCATC----- 438
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 439 -----CCTCAATC-----ACTTTA----- 453
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 DB 454 -----CCGAGACACACACGCAA-----GTGAAT 477
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
 DB 478 ATCGGT----- 483
 QY 194 SerTyrLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 DB 483 ----- 483
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 DB 484 -----GTGATGGGACG 495
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 DB 496 GTGACCGTGACT----- 507
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 507 ----- 507
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 507 ----- 507
 QY 294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 DB 507 ----- 507
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 507 ----- 507
 QY 334 MetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353

Db 507 ----- 507
 Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 507 ----- 507
 Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 507 ----- 507
 Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 508 ----- 508
 Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
 Db 547 ACTTGGCTAATTTGTCAATCCGCGGGGCTTCATTCTATGGGGGATAATTTGTTTCC 606
 Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 Db 607 ATCACCACGCTAGCGCGATGCGATTGTGGGCAACCGCGATTCTCAAGGCTTAGGCAAG 666
 Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 667 TTAAGGCAAGCTTTTGGAGCTTAGTAACGTGAGATTGGTAGAAGAAATGACAGATCTA 726
 Qy 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 Db 727 ATCACCCTCAAGGGCTTATGAAGCAATTTAAAGCAATTCAAACCGCTGATGCCATG 786
 Qy 494 LeuGlnLysAlaLeuGluLeuLysArg 502
 Db 787 CTCCAAACAGTCAATTCCTCCAAACGC 813

RESULT 30

ID AAT77463
 XX AAT77463 standard; DNA; 816 BP.

AC AAT77463;

DT 11-AUG-1997 (first entry)

DE H. pylori flagella associated protein ORF 29298130.aa.

KW Chronic gastritis; duodenal ulcer disease; activator; inhibitor;
 KW bacterial life cycle; vaccine; immunisation; detection; antisense;
 KW inhibition; flagella; flagellum; basal body; rod; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers
 CDS 1..816
 /*tag= a

FN WO9719098-A1.

PD 29-MAY-1997.

PF 15-NOV-1996; 96WO-US018542.

PR 17-NOV-1995; 95US-00561469.

PA (ASTR) ASTRA AB.

PI Smith DH;

DR WPI; 1997-298052/27.

DR P-PSDB; AAW24645.

XX Helicobacter pylori nucleic acid sequences and related proteins - used
 XX for diagnostics and therapeutics.

PS Claim 1; Page 105; 235pp; English.

XX The present sequence encodes a Helicobacter pylori flagella associated
 CC protein, which was found to be homologous to flagellar basal body rod
 CC protein following BLAST protein analysis. H. pylori has been strongly
 CC linked to chronic gastritis and duodenal ulcer disease. The nucleic acid
 CC sequences of the invention are used to evaluate compounds, especially
 CC activators or inhibitors of bacterial life cycle, for the ability to bind
 CC an H. pylori nucleic acid sequence. The nucleic acid sequences, and
 CC corresponding proteins, are also useful for generating vaccines for
 CC immunising subjects against H. pylori or for use in detecting the
 CC presence of Helicobacter species in a sample. Antisense nucleic acid
 CC sequences of these sequences are used to inhibit expression of a gene
 CC from Helicobacter species. H. pylori whole genomic DNA was isolated and
 CC nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-
 CC ended and ligated to unique BstXI-linker adapters in 100-1000 fold molar
 CC excess. These linkers are complementary to the BstXI-cut pMPX vectors,
 CC while the overhang is not self-complementary. Therefore the linkers will
 CC not concatamerise nor will the cut vector re-ligate itself easily. The
 CC linker-adaptor inserts were ligated to each of the 20 pMPX vectors to
 CC construct a series of shotgun subclone libraries. The purified DNA
 CC samples were then sequenced. Note: The ORF/protein reference number for
 CC this sequence was obtained from the related specification, WO9640893

SQ Sequence 816 BP; 237 A; 199 C; 163 G; 157 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,65e-14 Length: 816
 Score: 288.00 Matches: 107
 Percent Similarity: 28.68% Conservative: 39
 Best Local Similarity: 21.02% Mismatches: 109
 Query Match: 11.14% Indels: 254
 DB: 2 Gaps: 9

US-10-009-823A-1 (1-502) x AAT77463 (1-816)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 28 ATGTCCTCGCTCTCTCTATAGTCCCACTTCAGGATGTCGCCCAACACGACATTGAC 87
 Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 88 ACCACTTCAAAACAACATCGCCCAATGTCAATACCAACCGGGTTTAAAAAATCTCGCGGAT 147
 Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 Db 148 TTTAAGCACTGTTTTTACCAAGCGATGCAATACCGCGCACCAACACGACACACCACT 207
 Qy 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 Db 208 TTATCGCCAGATGCATGGAAGTGGCGCTTGGCTAGCCCTAGTGGCAATTACCAAAATG 267
 Qy 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
 Db 268 TTTTCGCAAGCGACGCTCTAAAGAAACGAGAGAAATTTAGATATTGCTATTACGGTAAA 327
 Qy 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 Db 328 GGCCTTTTTCAGTCCAGCTTCCTGATGGCACTACCGCTTACACAGGAGCGGGAATTC 387
 Qy 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 Db 388 AAGCTAGACGACGAGGCAATCTTGTAAACAGCGAGGCGCTATCTCCTCATC----- 438
 Qy 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db 439 -----CCTCAATC-----ACTTAA----- 453
 Qy 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 Db 454 -----CCGAAAGACACCAACGCA-----GTGAAT 477
 Qy 174 LeuGluAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193

16-JUN-1997 (first entry)

H. pylori flagella-associated membrane protein ORF 12ge20305orf11.

Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; envelope; membrane; flagella; ds.

Helicobacter pylori.

Key	Location/Qualifiers
CDS	1..837
FT	/*tag= a
FT	/note= "no stop codon given in sequence"

W09640893-A1.

19-DEC-1996.

06-JUN-1996; 96WO-US009122.

07-JUN-1995; 95US-00487032.

01-APR-1996; 96US-00630405.

(ASTR) ASTRA AB.

Smith D, Berglindh OT, Mellgaerd BL;

WPI: 1997-052306/05.

P-PSDB; AAW20863.

Helicobacter pylori nucleic acid sequences and related polypeptide(s) - useful for vaccines to treat or prevent H. pylori infection, and to detect Helicobacter.

Claim 1; Page 909; 1481pp; English.

The present sequence encodes a Helicobacter pylori flagella-associated membrane protein. The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors. The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts

Sequence 837 BP; 240 A; 201 C; 186 G; 210 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Length:
Score: 288.00	837
Percent Similarity: 28.68%	Matches: 107
Best Local Similarity: 21.02%	Conservative: 39
Query Match: 11.14%	Mismatches: 109
DB: 2	Indels: 254
	Gaps: 9

US-10-009-823A-1 (1-502) x AAT68116 (1-837)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20

Db 52 ATGCTCCCGCTCTCTCTATAGTGCCACTTCAGGATGTCGCCCAACAAACGCACATTGAC 111

Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40

Db 112 ACCACTTCAAAACACATGCCCAATGTCAATACCAACCGGTTTAAATAATCTCCGCGGAT 171

QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 172 TTTAAAGGACTGTTTACAGCGATGCAATACGCGGACCAACAAAGCAACACGACT 231
QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 232 TTATCGCCAGATGGCATGAAGTGGCGCTTGGCGTACGCCCTAGTGCATTACCAAAATG 291
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 292 TTTTCGCAAGGAGCCCTAAAGAAACGAGAGATAATTAGATATTGCTATTACAGGTAA 351
QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisThrThrArgAlaGlyAsnPhe 113
Db 352 GGCCTTTTCAAGTCCAGCTCTCTGATGCGACTACCGCTTACACAGAGCGGGAATTC 411
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 412 AAGCTAGACGACGCGGCAATCTGTACAGCGAGGCTATCTCTCATC----- 462
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 463 -----CCTCAATC-----ACTTTA----- 477
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 478 -----CCGAGACACACGCAA-----GTCAAT 501
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
Db 502 ATCGGT----- 507
QY 194 SerTyrLysGlyAsnGlyThrProPheThrSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 507 ----- 507
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 508 -----GTGATGGCAGC 519
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 520 GTGACCGTACT----- 531
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 531 ----- 531
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 531 ----- 531
QY 294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 531 ----- 531
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 531 ----- 531
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 531 ----- 531
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 531 ----- 531
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 531 ----- 531
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413

Db 532 -----CAAGGCTTGCAACGACTTCTTAAGCTGATCGGG-----CAAATC 570
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 571 ACTTGGCTAAATTTGTCAATCCGGGGGCTTCAATTCATGCGGGGATAATTTGTTTCC 630
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 631 ATCACCACGCTAGCGCGCATGCGATTGTTGGCAACCCGGATTCTCAAGGCTTAGGCAAG 690
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 691 TTAAGGCAAGGCTTTTGGAGCTTAGTACGTGAGATTGGTAGAAGAATGACATCTA 750
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 751 ATCAGCGCTCAAAGGCTTATGAAGCCAAATCTAAAGCAATTCAAACCGCTGATGCCATG 810
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 811 CTCCAACAGTCAATTCCTCCCAACGCG 837
RESULT 32
ADF03653
ID ADF03653 standard; DNA; 798 BP.
XX ADF03653;
AC ADF03653;
XX
DT 12-FEB-2004 (first entry)
XX
XX Bacterial polynucleotide #3938.
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant; gens; ds.
XX
XX Proteus mirabilis.
XX
XX US6605709-B1.
XX
PD 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2003-895291/82.
DR P-PSDB; ADF07825.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX
XX Disclosure; SEQ ID NO 3938; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides. Methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polynucleotide of the invention.
XX

SQ Sequence 798 BP; 233 A; 163 C; 189 G; 213 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,55e-14 Length: 798
 Score: 286.00 Matches: 100
 Percent Similarity: 30.57% Conservative: 55
 Best Local Similarity: 19.72% Mismatches: 98
 Query Match: 11.06% Indels: 254
 DB: 10 Gaps: 9

US-10-009-823A-1 (1-502) x ADF03653 (1-798)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 16 ATGATCGGTTCTTATGGATTGCTAAACACAGGGTGGATGCACAACTAACATGAT 75
 QY 21 ThrValSerAsnAlaAlaSerThrIleGlyTyrLysGlnGlnValVal 40
 DB 76 GTGATTCCAAACACTCCGCAACGTCAGCACAAATGGTTTAAAGCCAGCGTGGCGTT 135
 QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGlySer 56
 DB 136 TTGAGATTACTCTATCAACTATTCGTCACCGGAGCGATGACATCCGAGCAGCG 195
 QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 DB 196 AATGCGCCTTCTGTTTACAAATTGTTACTGTGTTCGCCCATGTCGCGACAGACGTTTA 255
 QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
 DB 256 CATAGCCAGGTAATTAGCCCAACTATGTTACCGGTGATGTTGCTATCAAGGCGAA 315
 QY 95 GlyPhePheGlnValThrLeuGluAspLysValHis---TyrThrArgAlaGlyAsnPhe 113
 DB 316 GGTITTTTCCATGTTCAATTACCTGATGTTGTCGCGATCTTATCTCGTGTGCTCTTTT 375
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 DB 376 CAAATGGACCAAAATGGCAACTAGTACCTCCAGTGGCTTTCAAAATCGTCCAGCG--- 432
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 433 -----ATTATTTTCCAGAAACCGCTAAAGGATGATGTA----- 468
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 DB 468 ----- 468
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
 DB 468 ----- 468
 QY 194 SerTrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 DB 468 ----- 468
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 DB 469 -----GGTGGTGGTGGTATTGTCAGTGTGAGATTGAGATCG 507
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 DB 508 CCTGCA----- 513
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 513 ----- 513
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 513 ----- 513
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313

DB 514 -----CCTCAA----- 519
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 519 ----- 519
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 DB 520 -----CAAGTAGGCGCAATTA--- 534
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
 DB 534 ----- 534
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 DB 535 -----ACCTCACC 543
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 DB 544 ACA----- 546
 QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
 DB 547 -----TTTATTATGATAGCGGTAGAAAGTGTGGGAAATCTGTACTTA 594
 QY 434 AlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 DB 595 GAAACAGCAGCTCCGGAGCACCACCTGAGAAATCGCGGTATTAACGGTGGCGGCTTG 654
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 DB 655 TTATATCAGGATATGTTGAAACCTCTAACGTTAATGTCGCGGAGAAATTCGTCATATG 714
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
 DB 715 ATCCAAACTCAGCGTGTATGAAATTAATAGTAAGCGATTTCACACTTCTGATCAGATG 774
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 DB 775 TTACAGAAACTCAGCAACTC 795

RESULT 33

AAZ88552 standard; DNA; 1800 BP.

AAZ88552;

04-MAY-2000 (first entry)

C. jejuni flgP and flgG DNA.

Basal body rod protein; flgP; flgG; flagellum; vaccine; immunogen;
 pathogenic bacteria; detection; antibacterial; ss.

Campylobacter jejuni.

Key Location/Qualifiers

CDS 101..913

FT /*tag= a

FT /product= "flgP"

FT 942..1733

FT /*tag= b

FT /product= "flgG"

PN US6020125-A.

PD 01-FEB-2000.

XX 07-JUN-1995; 95US-00483857.

XX 08-MAY-1995; 95US-00436748.

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS4197-AAS4564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1398 BP; 292 A; 401 C; 346 G; 359 T; 0 U; 0 Other;

Alignment Scores: 4.85e-13 Length: 1398
Pred. No.: 279.00 Matches: 77
Score: 49.77% Conservative: 30
Percent Similarity: 49.77% Mismatches: 92
Best Local Similarity: 35.81% Indels: 16
Query Match: 10.79% Gaps: 5
DB:

US-10-009-823A-1 (1-502) x AAS88418 (1-1398)
QY 285 PheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
Db 1386 TTTGTGAAGACCGGGAT-----AATAACTGGCAGGTCTACACCAGAT 1342
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlylleGlnProLeuThrLeu 324
Db 1341 AGCAGTGATCCAAACAGCATTTGCCAAG-----ACAGGACACCACTG 1300
QY 325 AspPheGlylleLysSerGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAla 344
Db 1299 GAATTT-----AATGCTAATGGCACATAGTGATGGCGATGGCGAATAAATATCGCA 1246
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 1245 ACCGGCGCAATTAACGGTGCAGAACCCGCCAGTTTAGTCTG-----AGTTCTCTCAAC 1192
QY 365 SerThrAlaArgAsn---GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrPro 383
Db 1191 TCCATGTCAGCAAAATACCGCGCTTAACAATATTGTGGCAACCCAGACGGCTACAAA 1132
QY 384 GlnGlyAspLeuValAspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSer 403
Db 1131 CCGGGCGATCTGGTGAGTTATCAATCAATGATGACGGTACGGTGTGCGCAACTATTTC 1072
QY 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
Db 1071 AACGAACAAACCCCACTCTGGGCGAGATTGTACTGGCGAACTTTGCCAACCAACAGGT 1012
QY 424 LeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Db 1011 CTGGCATCCGAAGGCGACACACGTCTGCTGGAGCGCAATCTTCTGGCGTGGCGCTGTG 952
QY 444 GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463
Db 951 GGGACAGCGCGACGGGAAACTTTGGCACCTTGACCAACCGGTGCGCTGGAAGCGTCCAAC 892
QY 464 ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483
Db 891 GTCGATCTCAGTACAGAACTGTGTCAATATGATGCTGCGGAGCTTAATCATCAGTCTAAC 832
QY 484 SerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeu 498
Db 831 GCCCAGACCATCAAAACCCAGGACCCAGATCTCTCAACACGCTGCTC 787

RESULT 35
AAV58977
ID AAV58977 standard; DNA; 1800 BP.
XX
AC AAV58977;
XX
DT 05-JAN-1999 (first entry)
XX
DE FlgFG operon.

QY 412 nileProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyLysAsnAsnHisTyr 432
Db 1460 AGTGGAGCTAGTTCAGTTTATAATCCAGCGGTCTTCATCTTATGGTGATAATCTTTA 1519
QY 432 rSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrG1 452
Db 1520 TCTTGAACAGGAGCAAGTGGTGCACCTTTTCGGGTATAGCAGCAAGATGGGCTTGG 1579
QY 452 YLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAs 472
Db 1580 ACAATTAACACATGGTTTATAGACTTATGTTTCAGCTTGTGAAGAAATCACACA 1639
QY 472 nMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspTh 492
Db 1640 TCTTATCACAGACAAAGAGCTTATGAGCGGGTTCATAGAGCCATTACAAACAGTATGA 1699
QY 492 rMetLeuGlnLysAlaLeuGluLysArg 502
Db 1700 TAGCTAGGAATGTAAATCAGCTTAAGCGA 1730

RESULT 34
AAS88418/c
ID AAS88418 standard; cDNA; 1398 BP.
XX
AC AAS88418;
XX
DT 13-FEB-2002, (first entry)
XX
DE DNA encoding novel human diagnostic protein #24222.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WC200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
PR (HYSE-) HYSEQ INC.
PA
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
PI P-PSDB; ABG24231.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 24222; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations

XX flgFG operon; flgF protein; flgG protein; basal body rod protein;
 KW flagellum; secretory diarrhoea; enteritis; vaccine; therapy; ds.
 XX

OS Campylobacter jejuni.

XX Key Location/Qualifiers

XX CDS 101..913

XX /*tag= a

XX /product= "flgF"

XX CDS 942..1733

XX /*tag= b

XX /product= "flgG"

PN US827654-A.

XX 27-OCT-1998.

XX 08-MAY-1995; 95US-00436748.

XX 08-MAY-1995; 95US-00436748.

XX (UTOR) UNIV TORONTO.

PA Louie H, Chan VL;

XX WPI; 1998-593983/50.

XX P-PSDB; AAW73072, AAW73073.

XX DNA encoding Campylobacter flagellum basal body rod proteins - useful for

XX recombinant production of the proteins for use as vaccines against the

XX bacterium, and for its detection, additionally with antibodies raised

XX with the protein.

XX Claim 1; Fig 1; 27pp; English.

XX This sequence represents the Campylobacter flgFG operon of the invention.

XX The encoded flgF and flgG proteins are basal body rod proteins of the

XX flagellum of the Campylobacter. The nucleic acid and host cell containing

XX it, are useful for the recombinant production of Campylobacter,

XX especially C. jejuni, basal body rod proteins. This bacterium is the

XX cause of secretory diarrhoea and enteritis. As such, the peptides

XX produced can be used to raise antibodies, which in turn can be used to

XX detect the presence of the organism diagnosis of the conditions.

XX Additionally, the peptides, and specifically the live vectors (e.g. pox-

XX or vaccinia virus) can be used as vaccines against the bacterium, and the

XX antibodies can be used for passive immunisation. The nucleic acids can

XX also be used to detect the presence of the bacterium

XX SQ Sequence 1800 BP; 637 A; 250 C; 364 G; 549 T; 0 U; 0 Other;

XX

XX Alignment Scores:

XX Pred. No.: 8.79e-13 Length: 1800

XX Score: 277.50 Matches: 132

XX Percent Similarity: 37.94% Conservative: 96

XX Best Local Similarity: 21.96% Mismatches: 217

XX Query Match: 10.73% Indels: 156

XX DB: 2 Gaps: 24

XX

XX US-10-009-823a-1 (1-502) x AAV58977 (1-1800)

XX

XX 1 MetMetGlySerLeuPheGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20

XX 101 ATGCAAAATCGATATTATCAAGCAACTGGCGGAATGTAACAGTTTAACTGAT 160

XX 21 ThrValSerAsnLeuAlaAsnAlaAsnThrIleGlyTyrGlnGlnValVal 40

XX 161 GTGATTACTAATATCTTGCATATATCAATACAAAGTGGATATAAAGAGATGATGTT 220

XX 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53

XX 221 ATTGCAGATTTTAAAGGATTTTAAAGAACTCAGGATGAGTTCCTATAGAAATCAC 280

XX

QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70
 Db 281 ACAAGAGATGCATCTCGTTTGTAAATACTACAATAGATGGAATCCCAAGTTTCTCAA 340
 QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
 Db 341 GAATATACGATTTTAGCTAGCTTCTTTAAAGGCCACAAACAATCCCTTGGATTGGCA 400
 QY 91 IleGlyGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysValHisTyrThr 108
 Db 401 ATGACTAGAGAAGATGCTTTTATTGTTGTCAGACCAAGAGTGGAGAAAGTAAAGATTACC 460
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 461 AAAGATGGAAATTTTCAACTTCATGATGAGGGTATTGTTGTAATAAGCAAGATACAG 520
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsn----- 139
 Db 521 GTATTAAAGTAGTGATTTATTTAATAATCTCAGAATGCTGCATACGCAATTCCTAATAGT 580
 QY 139 ----- 139
 Db 581 GCTGTTCAATTAGCGTTGATATAAAACGGAAGCATTTGAAGTTGATGGAGCTCAAAATGCA 640
 QY 140 -----IleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn----- 153
 Db 641 AGATTATTGTTAGCACAAAGTAGATGATATAAGAGCTTTGCAAAAAGATGGGATAATGTC 700
 QY 154 -----AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 Db 701 TATAAATAGATGATCTAACCCGATTAGAGATTGAAAACTCCCAATGCTATTTCGCCAA 760
 QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSer-----GluAla 185
 Db 761 GGTGTTTCTCAGGGATCAATGTTAATCCAGTTACTGAAATGGTAGGACTGATTGAAGCA 820
 QY 186 AsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThr 205
 Db 821 AAC-----AGAATGGTAGAAATGATCAAAAAGTTATGACAGCTCATATGGATGAC 871
 QY 206 SerAsnTyrSerTyrAlaGlnProMet----- 214
 Db 872 TTAATCAAGAGCTATCAATAGCTTGCAGCTTTAAATAATTTAAATAAATAAATAA 931
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
 Db 932 AGGATTAATAATGATGATGATCACTTCATAC-----TGCTGCTACAGGAATGGTAGCGCA 985
 QY 235 SerSerThrGly-SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspG1 254
 Db 986 GCAACACAAATTGATGTTTCTTCAATAACATCGCAATGTTAATACAGCA-----GG 1039
 QY 254 ySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeu-----MetSerG1 269
 Db 1040 TTTTAAGAAAAGTCGCGCAGAAATTTGCTGATCTTATGATCAAGTTATGAAGTATGCGAG 1099
 QY 269 yThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrG1 289
 Db 1100 AACTTCAACT-----TCAGTACTACTCTTCTCTCTCTCGGG 1135
 QY 289 ySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProG1 309
 Db 1136 TATAGAA----- 1142
 QY 309 nPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu----- 324
 Db 1143 -----GTGGGTGTGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 1186
 QY 325 -AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344
 Db 1187 TGAAGGAATTTAAATCAACAGTACT-----GATGGTCTTGATATGCTATTCGAGG 1240
 QY 344 alleGly-----ThrAspIleGlyLysLeuProSe 354

Db 1241 TAATGGTGTTCCTCAATACAACTTCCTGATGGCACTATAGAAATGGGCAATTTACAAA 1300
 QY 354 rMetMetProLleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerTh 374
 Db 1301 GATATG-----AGGATAATGAGGTAATATTGTA----- 1331
 QY 374 rArgArgTyrSerGlnAspGlyTyr-----ProGlnGlyAs 386
 Db 1332 -----AATTCAGATGGTTATAGACTTTTACCTGAAATGACAATACCTGAAGGC-- 1379
 QY 386 pLeuValAspValThrIleThrSerGluGly-----LysLeuGlnGlyLysTy 402
 Db 1380 -GCACAGCAATAATGTTGTACAGATGGAACTTCTGTATGCTACAGGGAGCA 1438
 QY 402 rSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAs 422
 Db 1439 ACAAGAACTCAAAAT-----GGCCAAAGTGGAGCTAGTTCAGTTTATAATCCAGC 1489
 QY 422 pGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProG1 442
 Db 1490 GGGTCTTCATTCATGGGTGATATCTTAACTTGAACAGAGCAAGTGGTGCACCTGT 1549
 QY 442 uPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSe 462
 Db 1550 TCGGGTATAGCAGGACCAAGATGGGCTTGGAACTATAGACATGGATTATAGAACTTAG 1609
 QY 462 rAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMe 482
 Db 1610 TAATGTTCACTGTGTGAGAAATCAGACATCTTATACAGACAAAGAGCTTATGAGC 1669
 QY 482 tAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysAr 502
 Db 1670 GGGTCTTAAGGCCATTACACAAAGTGATGATGCTAGGAATTTGAATCAGCTTAAGCG 1729
 QY 502 g 502
 Db 1730 A 1730
 RESULT 36
 AAZ88554
 ID AAZ88554 standard; DNA; 789 BP.
 XX
 AC AAZ88554;
 XX
 DT 04-MAY-2000 (first entry)
 XX
 DE C. jejuni flgG DNA.
 XX
 KW Basal body rod protein; flgG; flagellum; vaccine; immunogen;
 KW pathogenic bacteria; detection; antibacterial; ss.
 XX
 OS Campylobacter jejuni.
 XX
 PN US6020125-A.
 XX
 PD 01-FEB-2000.
 XX
 PF 07-JUN-1995; 95US-00483857.
 XX
 PR 08-MAY-1995; 95US-00436748.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Louie H, Chan VL;
 XX
 DR WPI; 2000-146875/13.
 DR P-PSDB; AA51377.
 XX
 PT Recombinant basal body rod protein producible by a transformed host for
 PT use in immunological vaccine compositions for in vivo administration to
 PT protect against diseases caused by bacterial pathogens.
 XX

Example 1; Col 21-22; 28pp; English.

This invention describes a novel recombinant basal body rod protein (I) producible by a transformed host containing an expression vector comprising a nucleic acid selected from; (a) an entire nucleic acid sequence of 1800 base pairs (bp), or the nucleic acid sequence of the flgG gene having 810 bp, both given in the specification; (b) a nucleic acid encoding the amino acid sequence of the flgG protein, a 270 residue sequence, given in the specification; (c) a nucleic acid sequence encoding a functional flgG basal body rod protein of a flagellum of a strain of Campylobacter; or (d) an immunogenic fragment of an flgG protein of (a), (b) or (c), and expression means operatively coupled to the nucleic acid molecule for expression by the host of a basal body rod protein of a flagellum of a strain of Campylobacter. (I) is useful in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens that produce basal body rod proteins. (I) are also useful as immunogens, as antigens in immunoassays, or for procedures for the detection of antibacterial, Campylobacter, basal body rod protein and/or peptide antibodies. The product of the invention has immunostimulatory activity. This sequence encodes the Campylobacter jejuni flgG protein described in the method of the invention

XX Sequence 789 BP; 260 A; 127 C; 179 G; 223 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,29e-10	Length:	789
Score:	237.50	Matches:	95
Percent Similarity:	27.45%	Conservative:	45
Best Local Similarity:	18.63%	Mismatches:	115
Query Match:	9.18%	Indels:	255
DB:	3	Gaps:	9

US-10-009-823A-1 (1-502) x AAZ88554 (1-789)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
Db	1	ATGATGAGATCACTTCATCTACTGCTGCTACAGGAATGTTAGCGCAGCAACACAAATGAT	60
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
Db	61	GTTACTTCAATAAATACATCGCCATGTTTAATACAGCAGGTTTAAAGAAAGTCGCGCGAA	120
QY	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
Db	121	TTTGCTGATCTATGATCAAGTTATGAAGTATGCAGGAATTCACCTTCAGCTACTACT	180
QY	61	GlnAlaGlyMetGlyAlaGlnValGly-----SerValArgThrIle	74
Db	181	CTTTCCTCTCGGTATAGAAAGTGGTGTGGTGTCCACAGCGGTAACTAAAGTT	240
QY	75	PheThrGlnGlyAlaPheGlu---ProGlyAsnSerValThrAspLeuAlaIleGlyGly	93
Db	241	TTTACTGAAGGAATTTAAATCAACAAAGTACTGATGCTCTTGATATGGCTATTCAGGT	300
QY	94	LysGlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsn	112
Db	301	AATGGGTTTTTCAATACAACTTCCTGATGGCACTATAGGATATAGATAAGAAATGGCAA	360
QY	113	PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySer	132
Db	361	TTTACAAAAGATAATGAAGGTATATATTGTAATTCAGATCGTTATAGACTT-----	411
QY	133	ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe	152
Db	411	-----	411
QY	153	AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal	172
Db	411	-----	411
QY	173	AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu	192

```
Db 411 ----- 411
Qy 193 GluSerTrpLysGlyAsnGlyThrProPheSerThrSerAsnTyrSerTyrAlaGln 212
Db 411 ----- 411
Qy 213 ProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGly 232
Db 411 ----- 411
Qy 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252
Db 411 ----- 411
Qy 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThr 272
Db 411 ----- 411
Qy 273 PheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292
Db 411 ----- 411
Qy 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312
Db 412 ----- 420
Qy 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
Db 421 ----- 438
Qy 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
Db 439 ----- 474
Qy 353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
Db 475 TCTGTATGCTACCA ----- 489
Qy 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
Db 489 ----- 489
Qy 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412
Db 490 ----- 519
Qy 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr 432
Db 520 GTGCGAGCTAGTTCAGTTATAATCCAGCGGCTTCATCTATGGTGATATCTTAT 579
Qy 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 580 CTTGAAACAGGAGCAAGTGTGCACCTGTTCGCGGTATACAGCAAGATGGCTTGA 639
Qy 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 640 ACAATAGACATGGATTATAGACTAGTATATGTTTCAGCTTGTGAAGAAATGACAGAT 699
Qy 473 MetIleIleLeuGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 700 CTTATCACAGCAAAAGAGCTTATAGCGGGTCTTAAGCCCATTAACACAGTGTATGAT 759
Qy 493 MetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 760 ATGCTAGGAATTGTAATCAGCTTACGCGA 789
```

RESULT 37

ABK74597

ID ABK74597 standard; DNA, 555 BP.

XX

AC ABK74597;

DT 13-AUG-2002 (first entry)

```
XX DE Bacillus licheniformis genomic sequence tag (GST) #1888.
XX KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX OS Bacillus licheniformis.
XX PN WO2002291113-A2.
XX PD 11-APR-2002.
XX PF 05-OCT-2001; 2001WO-US031437.
XX PR 06-OCT-2000; 2000US-00680598.
XX PR 27-MAR-2001; 2001US-0279526P.
XX PA (NOVO ) NOVOZYMES BIOTECH INC.
XX PA (NOVO ) NOVOZYMES AS.
XX PI Berka R, Clausen IG;
XX WIPI; 2002-416684/44.
XX PT Monitoring differential expression of several genes in first Bacillus
XX cell relative to expression of same genes in one or more second Bacillus
XX cells, by using substrate containing Bacillus genomic sequenced tag
XX array.
XX Claim 4; SEQ ID NO 1888; 200pp; English.
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions.
XX environmental stress or other physiological provocation. Extensive follow
XX up characterisation is unnecessary, when one spot on an array equals one
XX gene or one open reading frame, since sequence information is available.
XX This sequence represents a genomic sequence tag (GST) used in the method
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 555 BP; 163 A; 145 C; 129 G; 118 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	3.74e-10	Length:	555
Score:	237.00	Matches:	59
Percent Similarity:	49.70%	Conservative:	25
Best Local Similarity:	34.91%	Mismatches:	59
Query Match:	9.16%	Indels:	26
DB:	6	Gaps:	2

US-10-009-823A-1 (1-502) x ABK74597 (1-555)

Qy 360 ThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGln 379

Db 48 ACCATTACAGGAGAGCGGCTACTTCCGGTCCGATCAGGAGATGAAGTGTACTACAGA 107

Qy 380 AspGly-----TyrProGlnGlyAspLeuValAspVal----- 390

Db 108 TCAGGAAACTTCTACCGATCTCTGATGAAGTGTGACCTGCTACTGTGACGGCTGTTCGTC 167

QY 391 -----ThrileThr 393
Db 168 TTAAGCGCAATAATGGAAGATCAATTCGCGAAGACGCCCAAGCTTCAGCATCGCT 227
QY 394 SerGluGlyLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 228 CTGACGGACCGTTTCTTATGTCGATCAAAACAACAGACAGACCGCGCCGAGATC 287
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 288 TCGCTTGCACATTCAGCAACATCCGATTATCAAAAGCTGGAGACAACCTTTACCGC 347
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 348 GAAACATTAGTTCGGAGATCCGCAAGTCGTGTACTTGGGAGGCGCTCAGGAAA 407
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 408 ATTCAAACGAGTCGCTTGAATGTGCAACGTCGACCTTCAGAAAGAAATTTCCGAAATG 467
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 468 ATTATCGCGAGCGCGCTTCAGTCCATGCGAAATCATTAACACCTCTGTGAAT 527
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 528 CTTCAGGAACCTCGTCAATCTGAAGCGA 554

RESULT 38
ADM91627/c
ID ADM91627 standard; DNA; 742 BP.
XX AC ADM91627;
XX DT 15-JUL-2004 (first entry)
XX DE Lawsonia intracellularis specific nucleotide sequence SEQ ID NO:28.
XX KW Lawsonia intracellularis; infection; antibacterial; immunotherapy; gene;
XX LM ds; proliferative enteropathy.
XX OS Lawsonia intracellularis.
XX SS WO2004033631-A2.
XX PN 22-APR-2004.
XX PD 01-OCT-2003; 2003WO-US031318.
XX PF 04-OCT-2002; 2002US-0416395P.
XX PR (MINU) UNIV MINNESOTA.
XX PA Kapur V, Gebhart CU;
XX PI WPI; 2004-340902/31.
XX DR P-PSDB; ADM91689.

XX PT New nucleic acid that generates an amplification product from L.
XX PT intracellularis nucleic acid using an appropriate second nucleic acid
XX PT molecule, useful for treating and preventing L. intracellularis
XX PT infection.
XX PS Disclosure; SEQ ID NO 28; 87pp; English.
XX CC The present invention describes an isolated nucleic acid comprising a
XX CC nucleic acid molecule of at least 10 nucleotides in length having at
XX CC least 75% identity to SEQ ID NO:8741 (ADM91733), where any of the
XX CC molecule that is 10-29 nucleotides in length, under standard
XX CC amplification conditions, generates an amplification product from
XX CC Lawsonia intracellularis nucleic acid using an appropriate second nucleic
XX CC acid molecule. Also described: (1) a vector comprising the nucleic acid;

CC (2) a host cell comprising the vector; (3) an isolated polypeptide
CC encoded by the nucleic acid; (4) an article of manufacture comprising the
CC polypeptide; (5) an antibody having specific binding affinity for the
CC polypeptide; (6) a method for detecting the presence or absence of L.
CC intracellularis in a biological sample; (7) a method of preventing
CC infection by L. intracellularis in an animal; (8) a composition
CC comprising a first oligonucleotide primer and a second oligonucleotide
CC primer, where the first and second primers are each 10 to 50 nucleotides
CC in length, and where in the presence of L. intracellularis nucleic acid,
CC generate an amplification product under standard amplification
CC conditions, but do not generate an amplification product in the presence
CC of nucleic acid from an organism other than L. intracellularis; and (9)
CC an article of manufacture comprising the composition. The nucleic acid
CC sequence has antibacterial activity, and can be used in immunotherapy.
CC The nucleic acid and polypeptides are useful for treating and preventing
CC L. intracellularis infection e.g. proliferative enteropathy. The present
CC sequence represents an L. intracellularis specific nucleic acid molecule,
CC which is used in the exemplification of the present invention.

XX SQ Sequence 742 BP; 213 A; 149 C; 133 G; 245 T; 0 U; 2 Other;
Alignment Scores:
Pred. No.: 1.05e-08 Length: 742
Score: 221.00 Matches: 62
Percent Similarity: 48.97% Conservative: 33
Best Local Similarity: 31.96% Mismatches: 87
Query Match: 8.55% Indels: 12
DB: 12 Gaps: 3

US-10-009-823A-1 (1-502) x ADM91627 (1-742)
QY 316 GlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrp 335
Db 682 GGAAGTCGCTTACACCTCCATTAATTTTACCAGGGGACTCACAGAAC----- 632
QY 336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
Db 631 ACAGGTAATCTCTAAATATGGCTATTGAGGCAAGGGTTTTCACAGTA-----TTG 578
QY 356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
Db 577 GTTGTGTATCAATTAATGTATACAGACGACGCTTTTAACTTAATCAGATGGGACT 518
QY 376 ArgTyrSerGlnAspGlyTyrPro-----GlnGlyAspLeuVal 388
Db 517 CTTGTAAACAGCAATGATGATACCCCTACACAGAAATTTACAGTTCTCTCTGATCTAAA 458
QY 389 AspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
Db 457 TCTGTGCTATTCTTCAAGAGGCTCGTATTACTGTTTAGACGCAATGGACAGAAATC 398
QY 409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGly 428
Db 397 GCTGCTGGAGAAATACCTTTATATAGCTTTGTTAATCCAGCGGGACTGATGCAAAAGGA 338
QY 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
Db 337 CGTAATCTCTATATTTCCACAGAACATCCGCTGAAGCGGTAGAGGGTGTCCCTCGTGA 278
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Db 277 GAAACATTAGGACATATAGCTCAAGATCTCTTGAATGCTCTAATGTAGAGTTCTGTAT 218
QY 469 GluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db 217 GAAATGGTACAATGATTGTAGGACAAAGAGCATATGAATGAATTCAAAAGCAATCA 158
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 157 ACTTCAGCAACATGCTCTTAACTGCGATGCCCACTTAACGT 116
RESULT 39
ABK74555

QY 314 PheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyLeuSerGlnGlnAsn 333
DB 1039 TTTACTTCATCGGGA---GAACCGGAGGAGCTTTCTTGAATTTACCGGAGGAGAG--- 1092
QY 334 MetTrpAlaLapProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLeuPro 353
DB 1093 -----GCTGAGCTGCAAGGGAGCGCC-----GCCAAATCAAA 1128
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg-----AsnGlySer 371
DB 1129 GTAGCCGATGATCATGATTCAAAGGCGGATAAAGTCGCCGATCGCTGATGTGAA 1188
QY 372 SerSerThrArgGlySerGlnAspGlyTyrProGlnGlyAspLeuValAspValThr 391
DB 1189 GCGAGC-----GACAACGCCAACCGCAAAATCTCGTAAATGCTTC 1230
QY 392 IleThrSerGluGlyLeuGlnGlyLeuGlnGlyTyrSerAsnSerGlnValValAspPheTyr 411
DB 1231 ACCAAG-----AAATTCAAATCGGTGACAAACGACGACCGCTTTAGACTATTAC 1281
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlyLeuGlnGlyAsnAsnHis 431
DB 1282 -----GAGGCGATCATCGTGAATGCGGTAAAGCTCAGGAAACAAACAGA 1329
QY 432 TyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyr 451
DB 1330 CTGCCCAAAATACAGAAACG-----CTTGTAAACACTGCTGATTC 1371
QY 452 GlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetVal 471
DB 1372 AACAGACAGTCTGTT-----AGCGCCGTATCATTAGACGAAGAATGTCG 1416
QY 472 AsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAsp 491
DB 1417 AACATGATCCAAATTCAGATGCTTATACGCGGCTCGGAGATGATTACCTGTCAGGAC 1476
QY 492 ThrMetLeuGlnLysAlaLeu 498
DB 1477 GAAGTCTCGATAGATTAT 1497

RESULT 40
ADG32119
ID ADG32119 standard; DNA; 1432 BP.
AC AC
XX ADG32119;
XX DT 26-FEB-2004 (first entry)
XX DE DNA encoding a mutant B_licheniformis secreted polypeptide SeqID 89.
XX DE mutant; host cell; production yield; shelf life; product stability;
XX KW purity; secreted; gene; ds.
XX OS Synthetic.
XX OS Bacillus licheniformis.
XX FN WO2003093453-A2.
XX PD 13-NOV-2003.
XX PF 25-MAR-2003; 2003WO-DK000198.
XX PR 10-APR-2002; 2002DK-00000534.
XX PR (NOVO) NOVOZYMES AS.
XX PA Andersen JT, Jorgensen ST, Rasmussen MD, Olsen PB, Clausen IG;
XX WPI; 2004-053045/05.
XX DR P-PSDB; ADG32120.
XX PT New mutant Bacillus licheniformis host cell secreting 5 % less of one or

PT more secreted polypeptides than the parent host cell, useful for
PT producing a product of interest e.g. polypeptides, amino acids or
PT carbohydrates.
XX
PS Disclosure; SEQ ID NO 89; 422pp; English.
CC This invention relates to a novel Bacillus licheniformis (B.
CC licheniformis) mutant host cell derived from a parent B. licheniformis
CC host cell that is mutated in genes encoding secreted polypeptides.
CC Specifically, it refers to the generation of an improved Bacillus host
CC that reduces the need for product purification caused by contaminant
CC secreted native polypeptides in the culture medium. Accordingly, the
CC present invention describes reducing the expression of these native
CC proteins (e.g. proteolytic enzymes, nutrient uptake factors and signal
CC molecules), which in turn makes it easier to purify the heterologous
CC product of interest and therefore improving the production process.
CC Further benefits of a mutated host cell include an increase in total
CC production yield and a longevity of shelf life attributable to improved
CC product stability and purity. This polynucleotide is a DNA sequence
CC encoding a mutant B. licheniformis secreted polypeptide of the invention.
XX
SQ Sequence 1432 BP; 477 A; 287 C; 335 G; 333 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 9.82e-07 Length: 1432
Score: 201.00 Matches: 40
Percent Similarity: 66.25% Conservative: 13
Best Local Similarity: 50.00% Mismatches: 27
Query Match: 7.77% Indels: 0
DB: 12 Gaps: 0
US-10-009-823A-1 (1-502) x ADG32119 (1-1432)
QY 423 GlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGlu 442
DB 4 GGTATTATCAAAAGCTGGAGACAACTTTACCGCGAAACATTGAGTTCCGGAGATCGCAA 63
QY 443 PheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSer 462
DB 64 GTCGTTGTACTCGGGGAAGCGGCTCAGGAAAAAATTCAAACGAGTGCGCTTGAATGTCG 123
QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
DB 124 AACGTGACCTTTCAGAGAAATTTCCGAAATGATATTCGCGCAGCGGGCTTCAGTCC 183
QY 483 AsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLysArg 502
DB 184 AATGCGMAATCATTAACACCTCTGATGAATTTCTTCAAGAACTCGTCAATCTGAAGCGA 243

Search completed: November 1, 2004, 05:38:16
Job time : 912 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:35:12 ; Search time 195 Seconds

(without alignments)
531.115 Million cell updates/sec

Title: US-10-009-916A-1

Perfect score: 950

Sequence: 1 MKIKLFFVTISVITSLTISI.....DXPLPLGGGGARIACGVIPN 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476.5	50.2	178	2 Q7UCE0	Q7UCE0 shigella fl
2	476.5	50.2	190	2 Q83RB4	Q83RB4 shigella fl
3	474.5	49.9	173	1 SODC ECOLI	P53635 escherichia
4	473.5	49.8	190	2 Q8FH80	Q8FH80 escherichia
5	452	47.6	173	2 Q704S6	Q704S6 salmonella
6	452	47.6	173	2 CAFO6532	CAFO6532 salmonella
7	451	47.5	173	1 SODC_SALTY	Q88901 salmonella
8	449	47.3	173	2 Q8Z6P6	Q8Z6P6 salmonella
9	435.5	45.8	172	2 Q887F8	Q887F8 pseudomonas
10	425	44.7	177	1 SODI_SALTY	P53636 salmonella
11	425	44.7	177	2 Q704S7	Q704S7 salmonella
12	425	44.7	177	2 CAFO6531	CAFO6531 salmonella
13	420	44.2	174	1 SODC_BRUME	P58845 brucella me
14	417.5	43.9	174	2 Q6D531	Q6D531 erwinia car
15	414	43.6	171	2 Q9KRQ3	Q9KRQ3 vibrio chol
16	414	43.6	175	2 Q7AFX5	Q7AFX5 escherichia
17	414	43.6	274	2 Q8X9P0	Q8X9P0 escherichia
18	413.5	43.5	154	1 SODC_BRUAB	P15453 brucella ab
19	413.5	43.5	174	2 Q9MC02	Q9MC02 phage fels-
20	413.5	43.5	174	2 Q8ZQF7	Q8ZQF7 salmonella
21	413	43.5	175	2 Q7A6A1	Q7A6A1 escherichia
22	413	43.5	328	2 Q8X6B6	Q8X6B6 escherichia
23	411	43.3	170	2 Q7MPFM	Q7MPFM vibrio vuln
24	410	43.2	171	2 Q87G06	Q87G06 vibrio para
25	407	42.8	170	2 Q8D454	Q8D454 vibrio vuln
26	404	42.5	195	2 Q8D1A0	Q8D1A0 yersinia pe
27	404	42.5	201	2 Q8ZBN3	Q8ZBN3 yersinia pe
28	404	42.5	201	2 AAS60586	AAS60586 yersinia
29	399	42.0	170	2 Q7VWV9	Q7VWV9 bordetella
30	391	41.2	170	2 Q7WH53	Q7WH53 bordetella
31	391	41.2	173	1 SODC_PHOLE	P00446 photobacter

32 391 41.2 187 1 SODC_HAEPA P25842 haemophilus
33 391 41.2 194 2 Q7W9K6 Q7W9K6 bordetella
34 387 40.7 170 2 Q83AQ8 Q83AQ8 coxiella bu
35 387 40.7 186 1 SODC_NEIMA P57005 neisseria m
36 385 40.5 186 1 SODC_NEIMB Q59623 neisseria m
37 382 40.2 186 1 SODC_PASMU Q59689 pasteurella
38 376 39.6 190 1 SODC_ACTPL P24702 actinobacil
39 374 39.4 175 2 Q6G3C6 Q6G3C6 bartonella
40 374 39.4 199 1 SODC_HAEDU Q59452 haemophilus
41 373 39.3 187 1 SODC_HAEIN P25841 haemophilus
42 369 38.8 162 1 SODC_LEGPN P53637 legioretella
43 368 38.7 171 2 Q85021 Q59401 photobacter
44 365.5 38.5 185 1 SODC_FRATU Q59448 francisella
45 362 38.1 187 2 Q714V4 Q714V4 haemophilus

ALIGNMENTS

RESULT 1

Q7UCE0 PRELIMINARY; PRT; 178 AA.
ID Q7UCE0
AC Q7UCE0;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Superoxide dismutase (Cu-Zn)
GN Name=sodC; OrderedLocusNames=S1805;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24577;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 24577";
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2O)(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL; AE016983; AAP17140.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; SodCu; 1_CU_ZN.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 178 AA; 18167 MW; DC43B64712E8FA65 CRC64;

Query Match 50.2%; Score 476.5; DB 2; Length 178;

Best Local Similarity 54.1%; Pred. No. 6.5e-35;

Matches 92; Conservative 24; Mismatches 51; Indels 3; Gaps 2;

QY 10 SIVTISLTSTSVVLACSVTSEVHMDDNGIKQISGIVTFTDQKGLQIKTLKGLPAG 69

Db 10 SLAILALV--VATGAQAASEKVENMLVTSQGVGSGISGIVTITETDKGLERSPDLKALPPG 67

QY 70 BHGHFHEGSCGPAEHGHLTAGLQAHGYDPPKTKHKGPLNGHKGDLPRLVAVKADG 129

Db 68 BHGFHIAKSCQPAIKDGRKASAAESAGGHLDPQNTQKHGEGPETGHLGDLPALVNVNDG 127

QY 130 IAKETLLAPRL-TVKEIKGRWTMIHAGDNYSDKPLPLGGGGARIACGVI 178

```

Db      128 KATDAVIAPRLKSLDEIKDKALMVHVGGDNNSDQPKPLGGGGRYACGVI 177
      | : : : | | | | : : : | | | | | | | | | | | | | | | | | | |
RESULT 2
Q83RB4  PRELIMINARY;          PRT;      190 AA.
ID      Q83RB4
AC      Q83RB4
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Superoxide dismutase (Cu-Zn).
GN      Name=sodC; OrderedLocusNames=SFL673;
OS      Shigella flexneri.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Shigella.
OX      NCBI_TaxID=623;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=301 / Serotype 2a;
RX      MEDLINE=22272406; PubMed=12384590;
RA      Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA      Yang J., Yang F., Zhang J., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA      Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA      Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA      Yu J.;
RT      "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT      through comparison with genomes of Escherichia coli K12 and O157.";
RL      Nucleic Acids Res. 30:4432-4441(2002).
CC      -!- FUNCTION: Destroys radicals which are normally produced within the
CC      cells and which are toxic to biological systems (By similarity).
CC      -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC      -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR      EMBL; AE015188; AAM43254.1; -.
DR      HSSP; P53635; IESO.
DR      DR
DR      GO: GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR      GO: GO:0046872; F:metal ion binding; IEA.
DR      GO: GO:0006801; P:superoxide ion metabolism; IEA.
DR      InterPro: IPR001424; SOD_CU_ZN.
DR      Pfam: PF00080; SOD_CU_ZN; 1.
DR      ProDom: PD000465; SOD_CU_ZN; 1.
DR      PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW      Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ      SEQUENCE 190 AA; 19661 MW; 6217FD0BE596E253 CRC64;

Query Match          50.28; Score 476.5; DB 2; Length 190;
Best Local Similarity 54.18; Pred. No. 7e-35;
Matches 92; Conservative 24; Mismatches 51; Indels 3; Gaps 2;

Qy      10 SIVTISLTITSVVLACSVTSEVMIDNDIKQISGTVFTDTRKGLQIKTDLKLPG 69
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      22 SLATLALV--VATGAQAASEKVENLVITSGVGSIGSVTITETDKLGFSPDLKALPPG 79
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy      70 EHGPHIHGSCGPAEDHGLTACLOAHGHVDPDKTKHGKGLNGHKGDLRLVLVVKADG 129
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      80 EHGPHIHAKSCQATKDKGASAESAGGHLDPQNTKGKGGEGTGHGLDPLFALVWVNDG 139
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy      130 IAKETLLAPRL-TVKEIKGRVTMIVHAGGDNYSKPLPLGGGGARIACGVI 178
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      140 KATDAVIAPRLKSLDEIKDKALMVHVGGDNNSDQPKPLGGGGRYACGVI 189
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

RESULT 3
SODC_ECOLI
ID      SODC_ECOLI
AC      P53635; P96756;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-2004 (Rel. 45, Last annotation update)
DE      Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
DE      (Bacteriocuprein).

```

```

GN      Name=sodC; OrderedLocusNames=b1646, z2661, ECs2355;
OS      Escherichia coli, and
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562, 83334;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / W3110;
RX      MEDLINE=96196162; PubMed=8626323;
RA      Imlay K.R.C., Imlay J.A.;
RT      "Cloning and analysis of sodC, encoding the copper-zinc superoxide
RT      dismutase of Escherichia coli.";
RL      J. Bacteriol. 178:2564-2571(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX      MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA      Prostbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.;
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL      Nature 409:529-533(2001).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / RIMD 050952 / EHEC;
RX      MEDLINE=21156231; PubMed=11258796;
RA      Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA      Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT      "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.";
RL      DNA Res. 8:11-22(2001).
RN      [5]
RP      SEQUENCE OF 20-173 FROM N.A.
RC      STRAIN=QC871;
RX      MEDLINE=97156972; PubMed=9003353;
RA      Battistoni A., Folcarelli S., Gabbianelli R., Capo C., Rotilio G.;
RT      "The Cu,Zn superoxide dismutase from Escherichia coli retains
RT      monomeric structure at high protein concentration. Evidence for
RT      altered subunit interaction in all the bacteriocupreins.";
RL      Biochem. J. 320:713-716(1996).
RN      [6]
RP      SEQUENCE OF 20-36.
RX      MEDLINE=96196162; PubMed=8626323;
RA      Benov L.T., Fridovich I.;
RT      Unpublished results, cited by:
RL      Imlay K.R.C., Imlay J.A.;
RL      J. Bacteriol. 178:2564-2571(1996).
RN      [7]
RP      CHARACTERIZATION.
RX      MEDLINE=95014320; PubMed=7929223;
RA      Benov L.T., Fridovich I.;
RT      "Escherichia coli expresses a copper- and zinc-containing superoxide
RT      dismutase.";
RL      J. Biol. Chem. 269:25310-25314(1994).
RN      [8]
RP      CHARACTERIZATION.
RX      MEDLINE=95305591; PubMed=7786035;

```

RA Benov L.T., Chang L.Y., Day B., Fridovich I.;
 RT "Copper, zinc superoxide dismutase in *Escherichia coli*: periplasmic
 RT localization.";
 RL Arch. Biochem. Biophys. 319:508-511(1995).
 RN [9]
 RP CHARACTERIZATION.
 RX MEDLINE=96063713; PubMed=7589534;
 RA Battistoni A., Rotilio G.;
 RA "Isolation of an active and heat-stable monomeric form of Cu,Zn
 RT superoxide dismutase from the periplasmic space of *Escherichia coli*.";
 RL FEBS Lett. 374:199-202(1995).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 20-173.
 RX MEDLINE=98070606; PubMed=9405149;
 RA Pesce A., Capasso C., Battistoni A., Folcarelli S., Rotilio G.,
 RA Desideri A., Bolognesi M.;
 RA "Unique structural features of the monomeric Cu,Zn superoxide
 RT dismutase from *Escherichia coli*, revealed by X-ray crystallography.";
 RL J. Mol. Biol. 274:408-420(1997).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems. This enzyme is
 CC highly thermostable.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U51242; AAC03729.1; -;
 DR EMBL; AB000259; AAC74718.1; -;
 DR EMBL; AB005387; AAG56635.1; -;
 DR EMBL; AB002558; BAB35778.1; -;
 DR EMBL; X97766; CAA66363.1; -;
 DR FIR; C90923; C90923.
 DR FIR; G85771; G85771.
 DR FIR; JC6004; JC6004.
 DR PDB; 1ESO; X-ray; @=20-173.
 DR ECHOBASE; EB3195; -;
 DR Ecogen; EGI3419; sodC.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; Sod Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; FALSE_NEG.
 DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
 KW 3D-structure: Antioxidant; Complete proteome; Copper;
 KW Direct protein sequencing; Metal-binding; Oxidoreductase; Periplasmic;
 KW Signal; Zinc.
 FT CHAIN 1 19 Superoxide dismutase [Cu-Zn].
 FT METAL 20 173 Copper.
 FT METAL 67 67 Copper.
 FT METAL 69 69 Copper.
 FT METAL 92 92 Copper and zinc.
 FT METAL 101 101 Zinc.
 FT METAL 109 109 Zinc.
 FT METAL 112 112 Zinc.
 FT METAL 147 147 Copper.
 FT DISULFID 74 169
 FT STRAND 21 30
 FT TURN 31 32
 FT STRAND 33 46
 FT TURN 47 48
 FT STRAND 49 56
 FT STRAND 62 64
 FT STRAND 66 69
 FT STRAND 77 77

FT STRAND 84 84
 FT TURN 85 86
 FT HELIX 87 89
 FT STRAND 92 92
 FT TURN 95 96
 FT TURN 104 105
 FT TURN 110 111
 FT STRAND 112 112
 FT STRAND 116 118
 FT TURN 120 121
 FT STRAND 124 124
 FT STRAND 128 130
 FT TURN 131 132
 FT HELIX 136 139
 FT TURN 140 141
 FT STRAND 142 147
 FT HELIX 159 162
 FT STRAND 166 173
 SQ SEQUENCE 173 AA; 17681 MW; 9A0CB65F03AAB197 CRC64;
 Query Match 49.9%; Score 474.5; DB 1; Length 173;
 Best Local Similarity 54.1%; Pred. No. 9.5e-35;
 Matches 92; Conservative 24; Mismatches 51; Indels 3; Gaps 2;
 QY 10 SIVTISLTSTSVVLACSVTSEVHMIDDMGKIQSIGTFTDTDKGLQIKTKDLKGLPAG 69
 DB 5 SLATLALV--VATGAQAASEKVENLVTSQGVSGISGVITETDKGLFSPDLKALPPG 62
 QY 70 EHGPHIEGSGCGPAEHGHLTAGLQAHGYDDPKTKHKGPLNGHKGDLPLRVVVKADG 129
 DB 63 EHGPHIEGSGCGPATXDGKASAEAGAGHLDPQNTKGHEGPEGAGHLGDLPALVWVNDG 122
 QY 130 IAKETTLAPRL-TVKEIKGRVTMHHAGDNDVSDKPLPLGGGGARIACGVI 178
 DB 123 KATDAVIAPRLKSLDEIKDALMVHVGDNDMSQPKPLGGGGERYACGVI 172
 RESULT 4
 Q8F80 PRELIMINARY; PRT; 190 AA.
 ID Q8F80
 AC Q8F80;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Superoxide dismutase (EC 1.15.1.1).
 GN Name=sodC; OrderedLocNames=c2038;
 OS *Escherichia coli* O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CPT073 / ATCC 700928 / UPEC;
 EX MEDLINE=23388234; PubMed=12471157;
 RA Welch R.A., Surland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AE016761; AAN80498.1; -;
 DR HSSP; P53635; IESG.
 DR GO; GO:0004785; P:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; P:metal ion binding; IEA.
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.

DT	01-WAR-2002 (TReMBLrel. 28, Last sequence update)	
DT	01-OCT-2004 (TReMBLrel. 28, Last annotation update)	
DE	Copper-zinc superoxide dismutase.	
GN	Name=sodC; OrderedLocusNames=STY11692, t1308;	
OS	Salmonella typhi.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Salmonella.	
OX	NCBI_TaxID=601;	
LN	[1]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=CT18;	
RC	MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;	
RC	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,	
RA	Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,	
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,	
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,	
RA	Feltwell T., Hamlin N., Haque A., Hien T.F., Holroyd S., Jagels K.,	
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,	
RA	Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,	
RA	Whitehead S., Barrrell B.G.;	
RT	"Complete genome sequence of a multiple drug resistant Salmonella	
RT	enterica serovar Typhi CT18.";	
RL	Nature 413:848-852(2001).	
RL	[2]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=Ty2 / ATCC 700931;	
RC	MEDLINE=225311367; PubMed=12644504;	
RC	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,	
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;	
RA	"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2	
RT	and CT18.";	
RL	J. Bacteriol. 185:2330-2337(2003).	
CC	-I- FUNCTION: Destroys radicals which are normally produced within the	
CC	cells and which are toxic to biological systems (By similarity).	
CC	-I- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).	
CC	-I- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By	
CC	similarity).	
CC	-I- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.	
DR	EMBL; AL627271; CAD01927.1; -.	
DR	EMBL; AF016838; AAC68958.1; -.	
DR	HSSP; P53635; 1ESO.	
DR	GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.	
DR	GO; GO:0046972; F:metal ion binding; IEA.	
DR	GO; GO:0006801; P:superoxide metabolism; IEA.	
DR	InterPro; IPR001424; SOD_CU_ZN.	
DR	Pfam; PF00080; SOD_Cu; 1.	
DR	ProDom; PD000469; SOD_CU_ZN; 1.	
DR	PROSITE; PS00332; SOD_CU_ZN; 2; 1.	
KW	Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.	
SEQ	SEQUENCE 173 AA; 17735 MW, 00A8B86F25B4EF CRC64;	
Qy	Query Match 47.3%; Score 449; DB 2; Length 173;	
Db	Best Local Similarity 52.3%; Pred. No. 1.9e-32;	
Db	Matches 91; Conservative 22; Mismatches 49; Indels 12; Gaps 3;	
Qy	12 VTISLTSITSVVLAC-----SVTSEVHMIDNGIKSQISGTVTFTDQKGLQIKTDLKG 65	
Db	4 LSLAMVT-----LLACAGAAAEKVMNLUVTAGVCQSIGTVVIDTEGGLKFTPLKA 58	
Qy	66 LPAGEHGFIHIEGSCGPAEHDGHLTAGLOAHGYDPTDKGHEGPIGNHGKGLPLRV 125	
Db	59 LPPEHGFIHANGSCQPAIKDGKAVAAEAAGHGLDPQNTKGHEGSPGQGHGLGPLVW 118	
Qy	126 KADGIATKELLAPRLTVKELKGTWMIHAGGDNYSKPLPLGGGGARIACGVI 178	
Db	119 NNDGIATEPTAPRLKSLDELVKOKALMIHVGGDNMSQPKPLGGGGTRYACGVI 172	
RESULT 9		
Q887F8	PRELIMINARY;	
ID	Q887F8	PRT; 172 AA.
AC	Q887F8;	
DT	01-JUN-2003 (TReMBLrel. 24, Created)	

DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SOD_CU_1.
DR ProDom: PD000469; SOD_CU_ZN_1; FALSE_NEG.
DR PROSITE: PS00087; SOD_CU_ZN_1; FALSE_NEG.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW 3D-structure; Antioxidant; Complete proteome; Copper; Metal-binding;
KW Oxidoreductase; Periplasmic; Signal; Zinc.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 177 Superoxide dismutase [Cu-Zn] 1.
FT METAL 69 69 Copper.
FT METAL 71 71 Copper.
FT METAL 94 94 Copper and zinc.
FT METAL 103 103 Zinc.
FT METAL 112 112 Zinc.
FT METAL 115 115 Zinc.
FT METAL 150 150 Copper.
FT DISULFID 76 172 P -> T (in Ref. 2).
FT CONFLICT 49 49 M -> T (in Ref. 2 and 4).
FT CONFLICT 148 148
FT STRAND 23 31
FT STRAND 36 48
FT TURN 49 50
FT STRAND 51 58
FT STRAND 64 66
FT STRAND 68 71
FT STRAND 79 80
FT STRAND 85 86
FT TURN 87 88
FT HELIX 89 91
FT STRAND 94 94
FT TURN 97 98
FT TURN 106 107
FT TURN 109 110
FT TURN 113 114
FT STRAND 115 115
FT STRAND 119 121
FT TURN 123 124
FT STRAND 127 127
FT STRAND 131 133
FT TURN 134 135
FT TURN 139 142
FT HELIX 143 144
FT STRAND 146 150
FT TURN 162 165
FT STRAND 169 174
SQ SEQUENCE 177 AA; 18370 MW; 1EC743EE2AB38CAE CRC64;

Query Match 44.7%; Score 425; DB 1; Length 177;
Best Local Similarity 51.2%; Pred. No. 2.8e-30;
Matches 88; Conservative 22; Mismatches 54; Indels 8; Gaps 4;

QY 15 SLLTSITSVVLACSVTSE----VHMID--DNGIKQSIGTFTVTDKGLQIKTDLKLPLA 68
Db 4 TILSLVAGALISCSAMAENTLTVMNDALSSGTGENIGETVSETPYGLLFTPHLGLTP 63
QY 69 GEHGFHIEGSGCPAEHDGHLTAGLQAHGHYDPDKTKHKGPLGN-GHKGDLPRLVVKA 127
Db 64 GIHGFHVHTNPSCMPGKMGKEVPALMAGGHLDPKTKHGLGPNYDKHGLDLPGLVVNA 123
QY 128 DGIAKETILLAPRL-TVKEIKGRTVMHAGDNYSDKPLPLGGGARIACGVI 178
Db 124 DGTATYPLLAPRLKSLSELKSHLSLMIHKGDNYSKDPAPLGGGARPACGVI 175

RESULT 11
Q704S7 ID Q704S7 PRELIMINARY; PRT; 177 AA.
AC Q704S7
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Superoxide dismutase precursor (BC 1.15.1.1).
GN Name=sodC1;

OS Salmonella enterica subsp. enterica serovar Choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=119912;
RN [1]
RP SEQUENCE FROM N.A.
RA Amendola S., Ajello M., Pasquali P., Kroll J.S., Langford P.R.,
RA Rotilio G., Valenti P., Battistoni A.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: AJ620903; CAF06531.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0004784; F:superoxide dismutase activity; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; Sod Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN_1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Signal; Zinc.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 177 superoxide dismutase [Cu-Zn].
FT SEQUENCE 177 AA; 18370 MW; 1EC743EE2AB38CAE CRC64;

Query Match 44.7%; Score 425; DB 2; Length 177;
Best Local Similarity 51.2%; Pred. No. 2.8e-30;
Matches 88; Conservative 22; Mismatches 54; Indels 8; Gaps 4;

QY 15 SLLTSITSVVLACSVTSE----VHMID--DNGIKQSIGTFTVTDKGLQIKTDLKLPLA 68
Db 4 TILSLVAGALISCSAMAENTLTVMNDALSSGTGENIGETVSETPYGLLFTPHLGLTP 63
QY 69 GEHGFHIEGSGCPAEHDGHLTAGLQAHGHYDPDKTKHKGPLGN-GHKGDLPRLVVKA 127
Db 64 GIHGFHVHTNPSCMPGKMGKEVPALMAGGHLDPKTKHGLGPNYDKHGLDLPGLVVNA 123
QY 128 DGIAKETILLAPRL-TVKEIKGRTVMHAGDNYSDKPLPLGGGARIACGVI 178
Db 124 DGTATYPLLAPRLKSLSELKSHLSLMIHKGDNYSKDPAPLGGGARPACGVI 175

RESULT 12
CAF06531 ID CAF06531 PRELIMINARY; PRT; 177 AA.
AC CAF06531;
DT 10-MAY-2004 (TrEMBLrel. 27, Created)
DT 10-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Superoxide dismutase precursor (EC 1.15.1.1).
GN SODC1.
OS Salmonella enterica subsp. enterica serovar Choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella, Salmonella enterica.
OX NCBI_TaxID=119912;
RN [1]
RP SEQUENCE FROM N.A.
RA Amendola S., Ajello M., Pasquali P., Kroll J.S., Langford P.R.,
RA Rotilio G., Valenti P., Battistoni A.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: AJ620903; CAF06531.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0004784; F:superoxide dismutase activity; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; Sod Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN_1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Signal; Zinc.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 177 superoxide dismutase [Cu-Zn].
FT SEQUENCE 177 AA; 18370 MW; 1EC743EE2AB38CAE CRC64;

Query Match 44.7%; Score 425; DB 2; Length 177;
Best Local Similarity 51.2%; Pred. No. 2.8e-30;
Matches 88; Conservative 22; Mismatches 54; Indels 8; Gaps 4;

QY 15 SLLTSITSVVLACSVTSE----VHMID--DNGIKQSIGTFTVTDKGLQIKTDLKLPLA 68
Db 4 TILSLVAGALISCSAMAENTLTVMNDALSSGTGENIGETVSETPYGLLFTPHLGLTP 63
QY 69 GEHGFHIEGSGCPAEHDGHLTAGLQAHGHYDPDKTKHKGPLGN-GHKGDLPRLVVKA 127
Db 64 GIHGFHVHTNPSCMPGKMGKEVPALMAGGHLDPKTKHGLGPNYDKHGLDLPGLVVNA 123
QY 128 DGIAKETILLAPRL-TVKEIKGRTVMHAGDNYSDKPLPLGGGARIACGVI 178
Db 124 DGTATYPLLAPRLKSLSELKSHLSLMIHKGDNYSKDPAPLGGGARPACGVI 175

RESULT 12
CAF06531 ID CAF06531 PRELIMINARY; PRT; 177 AA.
AC CAF06531;
DT 10-MAY-2004 (TrEMBLrel. 27, Created)
DT 10-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Superoxide dismutase precursor (EC 1.15.1.1).
GN SODC1.
OS Salmonella enterica subsp. enterica serovar Choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella, Salmonella enterica.
OX NCBI_TaxID=119912;
RN [1]
RP SEQUENCE FROM N.A.
RA Amendola S., Ajello M., Pasquali P., Kroll J.S., Langford P.R.,
RA Rotilio G., Valenti P., Battistoni A.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: AJ620903; CAF06531.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0004784; F:superoxide dismutase activity; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; Sod Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN_1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Signal; Zinc.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 177 superoxide dismutase [Cu-Zn].
FT SEQUENCE 177 AA; 18370 MW; 1EC743EE2AB38CAE CRC64;

Query Match 44.7%; Score 425; DB 2; Length 177;
Best Local Similarity 51.2%; Pred. No. 2.8e-30;
Matches 88; Conservative 22; Mismatches 54; Indels 8; Gaps 4;

QY 15 SLLTSITSVVLACSVTSE----VHMID--DNGIKQSIGTFTVTDKGLQIKTDLKLPLA 68
Db 4 TILSLVAGALISCSAMAENTLTVMNDALSSGTGENIGETVSETPYGLLFTPHLGLTP 63
QY 69 GEHGFHIEGSGCPAEHDGHLTAGLQAHGHYDPDKTKHKGPLGN-GHKGDLPRLVVKA 127
Db 64 GIHGFHVHTNPSCMPGKMGKEVPALMAGGHLDPKTKHGLGPNYDKHGLDLPGLVVNA 123
QY 128 DGIAKETILLAPRL-TVKEIKGRTVMHAGDNYSDKPLPLGGGARIACGVI 178
Db 124 DGTATYPLLAPRLKSLSELKSHLSLMIHKGDNYSKDPAPLGGGARPACGVI 175


```

Db      1 MKLKALILATGVF-----FSCSTLAASVTTLNEALPTGAGSSIGDVSITPTPYGLLFT 53
QY      61 TDLKGLPAGEHGFHIEHGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEGPLGN-GHKGD 119
Db      54 PNDKLEAGLHGHVHENSCEPAEQDQKSVFALAGGHFDPKTKGHLGPDYDQGHVD 113
QY      120 LPRLVVKADGIKAKETLLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIA 178
Db      114 LPLGLVANADGTSTYPIVLPRIKSLFVQKHALMVHVGDNADMEAKLGGGGARMACGVI 173

RESULT 15
ID      Q9KRQ3 PRELIMINARY; PRT; 171 AA.
AC      Q9KRQ3;
DT      01-OCT-2000 (T-EMBLrel. 15, Created)
DT      01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT      01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE      Superoxide dismutase, Cu-Zn.
GN      OrderedLocusNames=VC1583;
OS      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
OX      NCBI_TaxID=666;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=El Tor N1691 / Serotype O1;
RX      MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA      Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA      Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA      Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA      Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA      Sellers P., McDonald L.A., Uitterback L.R., Fleischmann R.D.,
RA      Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA      Mekalanos J.J., Venter J.C., Fraser C.M.;
RT      "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT      cholerae.";
RL      Nature 406:477-483 (2000).
CC      -!- FUNCTION: Destroys radicals which are normally produced within the
CC      cells and which are toxic to biological systems (By similarity).
CC      -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC      -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR      EMBL; AE004235; AAF94737.1; -.
DR      FIC; A82183; A82183.
DR      HSSP; P00446; 1YAI.
DR      TIGR; VC1583; -.
DR      GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR      GO; GO:0046872; F:metal ion binding; IEA.
DR      GO; GO:0006801; P:superoxide metabolism; IEA.
DR      InterPro; IPR001424; SOD_CU_ZN.
DR      Pfam; PF00080; SOD_Cu; 1.
DR      ProDom; PD000469; SOD_CU_ZN; 1.
DR      PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR      Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
KW      SEQUENCE 171 AA; 17697 MW; 7483250CEA266C79 CRC64;
SQ

Query Match 43.6%; Score 414; DB 2; Length 171;
-Best Local Similarity 48.6%; Pred. No. 2.7e-29;
Matches 85; Conservative 19; Mismatches 59; Indels 12; Gaps 2;

QY      5 LFFVTISVTISLTISITSVLACSVTSEVHMIDNGIKQSIGTFTDTPDKGLQIKTDLK 64
Db      6 LFAVAFSSVSLAEMIVMT-----DLSSGQSVGTVTVDSEYGTFTPTLT 54
QY      65 GLPAGEHGFHIEHGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEGPLGN-GHKGDLPRL 123
Db      55 GLPAGLHGFHVGANGSCETSSKDGKTVLGGAGGHYDPQNTGKRGYPMTDNNHGLDLPAL 114
QY      124 VVKADGIKAKETLLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIA 178
Db      115 YVDAQQANQPVILASRFMAEVKGMALMVHAGDNNHSDHPNPLGGGGARIVCGVI 169

```

```

RESULT 16
ID      Q7AFX5 PRELIMINARY; PRT; 175 AA.
AC      Q7AFX5;
DT      05-JUL-2004 (T-EMBLrel. 27, Created)
DT      05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE      Putative copper/zinc-superoxide dismutase.
GN      OrderedLocusNames=ECs1120;
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=83334;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX      MEDLINE=21156231; PubMed=11258796;
RA      Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA      Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT      "Complete genome sequence of enterohemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.";
RL      DNA Res. 8:11-22 (2001).
CC      -!- FUNCTION: Destroys radicals which are normally produced within the
CC      cells and which are toxic to biological systems (By similarity).
CC      -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC      -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR      EMBL; AP002554; BAB34543.1; -.
DR      InterPro; IPR001424; SOD_CU_ZN.
DR      Pfam; PF00080; SOD_Cu; 1.
DR      ProDom; PD000469; SOD_CU_ZN; 1.
DR      PROSITE; PS00332; SOD_CU_ZN_2; 1.
DR      Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
KW      SEQUENCE 175 AA; 18259 MW; 00D8A6F5B7B94792 CRC64;
SQ

Query Match 43.6%; Score 414; DB 2; Length 175;
-Best Local Similarity 47.8%; Pred. No. 2.7e-29;
Matches 88; Conservative 27; Mismatches 53; Indels 15; Gaps 5;

QY      1 MKIKLFFVTISVTISLTISITSVLACSVTSEVHMIDNGIKQSIGTFTDTPDKG 56
Db      1 MKCKT-----IAAIAMLTA-----ASCGYAAQEVPMNLVSADGKEVSGIKTIQETPYG 50
QY      57 LQIKTDLKLPAEGHGFHIEHGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEPLG-NG 115
Db      51 LLFTPALHSLSEGIHGFHVGKNCAPALKDQKPVAAALSAGHFDPKTKGHLGFWSPDG 110
QY      116 HKGDLPLRVVKADGIKAKETLLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIA 174
Db      111 HPGDLPALFVTHDGKANYVPLAPRLNLSLKEIKGRSLMLHAGDNNHHDHPEPLGGGGARMA 170
QY      175 CGVI 178
Db      171 CGII 174

RESULT 17
ID      Q8X9P0 PRELIMINARY; PRT; 274 AA.
AC      Q8X9P0;
DT      01-MAR-2002 (T-EMBLrel. 20, Created)
DT      01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT      01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE      Putative copper-zinc superoxide dismutase encoded within prophage CP-
DE      933R.
GN      OrderedLocusNames=z2347;
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

OC Enterobacteriaceae; Escherichia.
 RN [1]_TaxID=83334;
 RP SEQUENCE FROM N.A.
 RC STRAIN=015:H7 / EDL933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Ferris N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamoukis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533 (2001).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AF005367; AAG56394.1; -.
 DR PIR; F85741; F85741.
 DR PIR; H90768; H90768.
 DR HSSP; P53636; 1EQW.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; Sod Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 1.
 KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
 SQ SEQUENCE 274 AA; 28777 MW; 51DF1AFBF50989 CRC64;
 Query Match 43.6%; Score 414; DB 2; Length 274;
 Best Local Similarity 47.8%; Pred. No. 4.6e-29;
 Matches 88; Conservative 27; Mismatches 53; Indels 16; Gaps 5;
 QY 1 MKIKLFFVTSIVTSLTSITSVLACSVTSE-----VHMDDNGIKOSIGTIVTFTDCKG 56
 Db 100 MKKCI-----IAAIAMLTA-----ASCCYAAEQVPMNLVSADGKSVISGKITIQETPYG 149
 QY 57 LQIKTDKGLPAGEHGFHIEGGSCGPAEHDGHLTAGLQAHGHYDPDKTGKHEGPG 115
 Db 150 LLFTPALHSLSEGIHGFVHEKNCALXGKPVAAISAGHFDPKNTGKHLGPSWPDG 209
 QY 116 HKGDLPRLVVADGIKETLLAPRL-TVKEIKGRTVMHAGGDNYSKPLPLGGGGARIA 174
 Db 210 HPGDLPALFVTHDGKANYPVLAAPRLNSLKEIKGRSLMLHAGGDNHHDHPEPLGGGGARIA 269
 QY 175 CGVI 178
 Db 270 CGII 273
 RESULT 18
 SODC BRUAB STANDARD; PRT; 154 AA.
 AC P15453;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 05-JUN-2004 (Rel. 44, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
 GN Name=sodC;
 OS Brucella abortus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OC [1]
 RN NCBI_TaxID=235;
 RP SEQUENCE.
 RX MEDLINE=90148961; PubMed=2105741;
 RA Beck B.L., Tabatabai L.B., Mayfield J.E.;
 RT "A protein isolated from Brucella abortus is a Cu-Zn superoxide

dismutase.";
 RL Biochemistry 29:372-376 (1990).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR PIR; A33893; A33893.
 DR HSSP; P53636; 1EQW.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; Sod Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 1.
 KW Antioxidant; Copper; Direct protein sequencing; Metal-binding;
 KW Oxidoreductase; Periplasmic; Zinc.
 FT METAL 48 48 Copper (By similarity).
 FT METAL 50 50 Copper (By similarity).
 FT METAL 73 73 Copper and zinc (By similarity).
 FT METAL 82 82 Zinc (By similarity).
 FT METAL 90 90 Zinc (By similarity).
 FT METAL 93 93 Zinc (By similarity).
 FT METAL 128 128 Copper (By similarity).
 FT DISULFID 55 150 By similarity.
 SQ SEQUENCE 154 AA; 16072 MW; 4672C31481704468 CRC64;
 Query Match 43.5%; Score 413.5; DB 1; Length 154;
 Best Local Similarity 52.0%; Pred. No. 2.6e-29;
 Matches 79; Conservative 19; Mismatches 53; Indels 1; Gaps 1;
 QY 28 SVTSEVHMDDNGIKOSIGTIVTFTDCKGLQIKTDKGLPAGEHGFHIEGGSCGPAEHD 87
 Db 2 STTVQMEALPTGPKCEVGTTVISEAPGLHFKVMKELTPYGHGFVHNPSCAPGEKD 61
 QY 88 GHLTAGLQAHGHYDPDKTGKHEGPGNGHKGDLPRLVVADGIKETLLAPRL-TVKEIK 146
 Db 62 GKIVPALAAGGHYDPGNTHHLGPEGDGHMGDLPRLSANADGKVSFTVVAHLKLAEIK 121
 QY 147 GRTVMHAGGDNYSKPLPLGGGGARIA 178
 Db 122 QRLMVHVGDNYSKPEPLGGGGARFACGVI 153
 RESULT 19
 Q9MCO2 PRELIMINARY; PRT; 174 AA.
 AC Q9MCO2;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Putative Cu/Zn superoxide dismutase.
 GN Name=sodCIII;
 OS Phage Fels-1.
 OC Viruses.
 OC NCBI_TaxID=128975;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21065122; PubMed=11136448;
 RA Figueroa-Bossi N., Uzau S., Maloriot D., Bossi L.;
 RT "Variable assortment of prophages provides a transferable repertoire
 RT of pathogenic determinants in Salmonella".
 RL Mol. Microbiol. 39:260-271 (2001).
 DR EMBL; AF254754; AAF82484.1; -.
 DR HSSP; P53636; 1EQW.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; Sod Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.

AC Q8X6B6;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Putative superoxide dismutase.
 GN OrderedLocusNames=z3312;
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 RN NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDJ933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.F., Potamocousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AE005440; AAG57201.1; -;
 DR PIR; E85842; E85842.
 DR PIR; E90877; E90877.
 DR HSP; P53636; 1EQW.
 DR GO; GO:004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0045872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 2; 1.
 KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
 SQ SEQUENCE 328 AA; 34217 MW; E3DE6E5C988E5A6 CRC64;
 Query Match 43.5%; Score 413; DB 2; Length 328;
 Best Local Similarity 47.8%; Pred. No. 6.9e-29;
 Matches 88; Conservative 27; Mismatches 53; Indels 16; Gaps 5;
 QY 1 MKIKLFFVTSIVTSLTSTSVVLAQSVTSE----VHMIDNDGKOSIGTVFTFTDQK 56
 DB 154 MKCKI-----IAAIAMLTA-----ASCGYAAEQEVPNNLVASDGKEVSIKTIQETPYG 203
 QY 57 LQIKTLKGLPAGBHGHIHEGSCGPAEDHGHITAGLQAHGHYDPKTKGHEGPIG-NG 115
 DB 204 LLFTPALHLSSEGHGHHVHEKNCAPALKDGFVAALSAGGHDFPNTKGHLGPNWSPDG 263
 QY 116 HKGDLPLRVKADGIKATLAPRL-TVSEIKGRVTMIHAGSDNYDKPLPLGGGGARIA 174
 DB 264 HLGDLPALFVTHDGKANYVPVLAPELNSLKEIKGRSLMLHAGDNDHHDHPEPLGGGGARWA 323
 QY 175 CGVI 178
 DB 324 CGII 327
 RESULT 23
 Q7MF9
 ID Q7MF9 PRELIMINARY; PRT; 170 AA.
 AC Q7MF9;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Superoxide dismutase, Cu-Zn.
 GN Name=VVA0291;
 OS Vibrio vulnificus (strain YJ016).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=196600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14656965;
 RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
 RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
 RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
 RT "Comparative genome analysis of Vibrio vulnificus, a marine
 RT pathogen.";
 RL Genome Res. 13:2577-2587(2003).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AF005345; BAC96317.1; -;
 DR GO; GO:004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0045872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN; 1; UNKNOWN 1.
 DR Copper; Metal-binding; Oxidoreductase; Zinc.
 KW SEQUENCE 170 AA; 17613 MW; 5AA2C93F1176704A CRC64;
 Query Match 43.3%; Score 411; DB 2; Length 170;
 Best Local Similarity 48.2%; Pred. No. 4.9e-29;
 Matches 82; Conservative 26; Mismatches 56; Indels 6; Gaps 2;
 QY 10 SIVTSLTSTSVVLAQSVTSEVHMIDNDGKOSIGTVFTDQKGLQIKTLKGLPAG 69
 DB 5 TLLAAILLYSTSSFAQSLVD-----MKDLSSNQLTGTVTSSSDYGTFTFDKGLPAG 59
 QY 70 EHGFIHEGSCGPAEDHGHITAGLQAHGHYDPKTKGHEGP-LGNHGKGLPLRVKAD 128
 DB 60 LHGFELHANGSCSSNKGDKVVLGGAGGHYDPQNSGKHGYPWTEDNHLGDLPALFVDAS 119
 QY 129 GKAKETLLAPRLTVKIKGRVTMIHAGSDNYDKPLPLGGGGARIA 178
 DB 120 GNASQVLAPELNSLKEIKGRSLMLHAGDNDHHDHPEPLGGGGARIVCGVI 169
 RESULT 24
 Q87G06
 ID Q87G06 PRELIMINARY; PRT; 171 AA.
 AC Q87G06;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Superoxide dismutase, Cu-Zn.
 GN OrderedLocusNames=VPA1514;
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nejima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae";
 RL Lancet 361:743-749(2003).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By


```

RESULT 27
Q8ZBN3 PRELIMINARY; PRT; 201 AA.
ID Q8ZBN3; Q74XS8;
AC AAS60586 PRELIMINARY; PRT; 201 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-VAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1)
GN Name=sodC; OrderedLocusNames=YPO311, YPO3375;
OS Versinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017128; AAS60586.1;
SQ SEQUENCE 201 AA; 20675 MW; 75D25ABE48BF8DCD CRC64;

Query Match 42.5%; Score 404; DB 2; Length 201;
Best Local Similarity 54.2%; Pred. No. 2.6e-28;
Matches 83; Conservative 18; Mismatches 50; Indels 2; Gaps 2;

QY 28 SVTSEVHMIDDNGIKQSIGVTFTDQKGLQIKTDKGLPAGEHGPHIHEGSCGPAEHD 87
DB 48 SMTVKINESLPQNGKALGTVTETAYGLTFPHLTGLAPGHGPHLHKPSCAPGMKD 107
QY 88 GHLTAGLQAHGYDPDKTGKEGFLGN-GHKGDPLRLVVKADGIAKLTLLAPRL-TVKEI 145
DB 108 GKAVPALAAGGHLDPNKTGVHLGPNYDKHGLGDLPLGVNADGTATYPVLAFLKSLSEV 167
QY 146 KGRVTMIHAGDNYSDKPLPLGGGGARACGVI 178
DB 168 KQALMIHAGDNYSDHPMPPLGGGGARMACGVI 200

RESULT 29
Q7VWVW9 PRELIMINARY; PRT; 170 AA.
ID Q7VWVW9
AC Q7VWVW9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative superoxide dismutase (EC 1.15.1.1).
GN OrderedLocusNames=BP2067;
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.B., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).

```

```
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -1- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: BX640417; CAE42345.1; -.
DR GO: GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006801; P:superoxide metabolism; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SOD_Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN; 1.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; UNKNOWN 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 170 AA; 17422 MW; BFBC899F119BSA5E CRC64;

Query Match 42.0%; Score 399; DB 2; Length 170;
Best Local Similarity 47.6%; Pred. No. 6e-28;
Matches 81; Conservative 25; Mismatches 60; Indels 4; Gaps 3;

QY 10 SIVTISLTSITSVVLACSVTSEVHMIDNGIKQSIGTFTTDTDKGLQIKTDLKGLPAG 69
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3 SLIALA-LAAMSGVAAQSV--EMSLSPKGDQKIGIKYEQTDYGVLFPTDLHGLAPG 59
QY 70 EHGFIHGGSCGPAEHGHLTAGLQAHGYDPDKTGKHEGPLG-NGHKGDLPRLVVKAD 128
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 60 VHGFLHDKPSCAMTKVDGKDMVGGAGGHWDPDKTGAHKGPYDKTGKGDLPVAVVAD 119

QY 129 GIAKETLLAPRLVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 120 GKATYPLAPRLKVLADLGHALMIHVGDNYSDHPAALGGGGGACGVM 169

RESULT 30
ID Q7WH53 PRELIMINARY; PRT; 170 AA.
AC Q7WH53;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative superoxide dismutase (EC 1.15.1.1).
GN OrderedLocusNames=BB3357;
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jegeris K.,
RA Leather S., Moule S., Norbertzak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -1- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: BX640447; CAE33849.1; -.
DR GO: GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
```

```
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006801; P:superoxide metabolism; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SOD_Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN; 1.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; UNKNOWN 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 170 AA; 17394 MW; DAE2CCB50C4FA5D CRC64;

Query Match 41.2%; Score 391; DB 2; Length 170;
Best Local Similarity 47.1%; Pred. No. 3.1e-27;
Matches 80; Conservative 24; Mismatches 62; Indels 4; Gaps 3;

QY 10 SIVTISLTSITSVVLACSVTSEVHMIDNGIKQSIGTFTTDTDKGLQIKTDLKGLPAG 69
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3 SLIALA-LAAMSGVAAQSV--EMSLSPKGDQKIGIKYEQTDYGVLFPTDLHGLAPG 59
QY 70 EHGFIHGGSCGPAEHGHLTAGLQAHGYDPDKTGKHEGPLG-NGHKGDLPRLVVKAD 128
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 60 VHGFLHDKPSCAMTKVDGKDMVGGAGGHWDPDKTGAHKGPYDKTGKGDLPVAVVAD 119

QY 129 GIAKETLLAPRLVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 120 GKATYPLAPRLKVLADLGHALMIHVGDNYSDHPAALGGGGGACGVM 169

RESULT 31
ID SODC_PHOLE STANDARD; PRT; 173 AA.
AC P00446;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN Name=sodC;
OS Photobacterium leiognathi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=658;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87109348; PubMed=3805055;
RA Steinman H.M.;
RT "Bacteriocuprein superoxide dismutase of Photobacterium leiognathi.
RT Isolation and sequence of the gene and evidence for a precursor
RT form.";
RL J. Biol. Chem. 262:1882-1887(1987).
RN [2]
RP SEQUENCE OF 23-173.
RX MEDLINE=83289129; PubMed=6884993;
RA Steffens G.J., Bannister J.V., Bannister W.H., Flohe L., Gunzler W.A.,
RA Kim S.-M.A., Otting F.;
RT "The primary structure of Cu-Zn superoxide dismutase from
RT Photobacterium leiognathi: evidence for a separate evolution of Cu-Zn
RT superoxide dismutase in bacteria.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:675-690(1983).
RN [3]
RP DISCUSSION OF POSSIBLE GENE TRANSFER FROM EUKARYOTES.
RX MEDLINE=85113139; PubMed=3855538;
RA Bannister J.V., Parker M.W.;
RT "The presence of a copper/zinc superoxide dismutase in the bacterium
RT Photobacterium leiognathi: a likely case of gene transfer from
RT eukaryotes to prokaryotes.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:149-152(1985).
RN [4]
RP DISCUSSION OF POSSIBLE GENE TRANSFER FROM EUKARYOTES.
RA Leunissen J.A.M., de Jong W.W.;
RT "Copper/zinc superoxide dismutase: how likely is gene transfer from
RT penicillium to Photobacterium leiognathi.";
RL J. Mol. Evol. 23:250-258(1986).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
```



```

DR EMBL; M84013; AAA24954.1; -.
DR PIR; B41654; B41654.
DR DR HSGP; P24702; 2APS.
DR DR InterPro; IPR001424; SOD_CU_ZN.
DR DR Pfam; PF00800; Sod_Cu; 1.
DR DR ProDom; PD000469; SOD_CU_ZN; 1.
DR DR PROSITE; PS00087; SOD_CU_ZN; 1.
DR DR PROSITE; PS00332; SOD_CU_ZN; 1.
DR KW Antioxidant; Copper; Metal-Binding; Oxidoreductase; Periplasmic;
Signal; Zinc.
FT SIGNAL 1 23 Potential.
FT FT CHAIN 24 187 Superoxide dismutase [Cu-Zn].
FT FT METAL 80 80 Copper (By similarity).
FT FT METAL 82 82 Copper (By similarity).
FT FT METAL 105 105 Copper and zinc (By similarity).
FT FT METAL 114 114 Zinc (By similarity).
FT FT METAL 123 123 Zinc (By similarity).
FT FT METAL 126 126 Zinc (By similarity).
FT FT METAL 161 161 Copper (By similarity).
FT FT DISULFID 87 183 By similarity.
FT SEQUENCE 187 AA; 19510 MW; AOC3A61EFAF201D5 CRC64;

Query Match 41.2%; Score 391; DB 1; Length 187;
Best Local Similarity 51.0%; Pred. No. 3.5e-27;
Matches 78; Conservative 22; Mismatches 51; Indels 2; Gaps 2;

QY 28 SVTSSVHMIDNGIKSQISGIVTFETDTKGLQIKIDLKGLPAGEHGPHIHEGSGCPAEHD 87
Db 34 STEVKQQLDPANGKNQGVTVITTESNYGLVFTPNLQGLAEGHGLPHIHENFSCDPKEKD 93
QY 88 GHLTAGLQAHGHYDDPKTKGHEGP-LGNHGKHGDLPLRVKADGIAKETLAPRL-TVKEI 145
Db 94 GKLTSGLAAGHWDPKGAQKHGYPQDDAHLGLDLPALTVLHDGTATNPVLAPRLKKLDEV 153
QY 146 KGRFTWITAGGDNYSKDLPLGGGGGARIAGVI 178
Db 154 RGHSMITHAGGDNHSDHPAPLGGGQPRWACGVI 186

RESULT 33
Q7R9K6 PRELIMINARY; PRT; 194 AA.
ID AC Q7R9K6 PRELIMINARY; PRT; 194 AA.
IC Q7R9K6;
AD Q7R9K6;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Putative superoxide dismutase (EC 1.15.1.1).
GN OrderedLocusNames=BpP1751;
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Acaligenaceae; Bordetella.
OC NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STEAN=12822 / ATCC BAA-587;
RX MEDLINE=22877954; PubMed=12910271; DOI=10.1038/ngl1227;
RA Parkhill J., Sebaghia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham A., Bason N., Cherevach I.,
RA Chillingworth T., Collins S., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goale A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
FT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003)".
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By

```

```

ENBL; AL162756; CAB84845.1; -.
PIR; E81855; E81855.
HSP; P24702, 2APS.
InterPro; IER001424; SOD_CU_ZN.
Fam; PF00080; SOD_Cu; 1.
ProDom; PD000469; SOD_CU_ZN; 1.
PROSITE; PS00087; SOD_CU_ZN 1; 1.
PROSITE; PS00332; SOD_CU_ZN 2; 1.

```

[2] SEQUENCE FROM N.A.
STRAIN=MC58 / Serogroup B;
MEDLINE=20175755; PubMed10710307;
Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
Nelson K.E., Eisen J.A., Karchum K.A., Hood D.W., Peden J.F.,
Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,
Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
Dourthy B.A., Mason T.N., Ciecko A., Parksey D.S., Blair E.,
Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
Qin H., Vamathevan J.J., Gill J., Scarlato V., Masignani V., Pizza M.,
Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
Venter J.C.;
"Complete genome sequence of *Neisseria meningitidis* serogroup B strain
MC58.";
Science 287:1809-1815 (2000) .
[3]
SEQUENCE OF 87-173 FROM N.A.
STRAIN=MC58 / Serogroup B;
MEDLINE=96118708; PubMed74596539;
Kroll J.S., Laceyford B., Willey J.P., Kroll J.D.

RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RL from the eukaryotic enzyme, and not so rare after all!";
RC
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001313; CAA04674.1; --
CC EMBL; AB002488; AAF41762.1; --
CC EMBL; X83126; CAA58207.1; --
CC F1R; F81088; F81088.
CC HSP; P24702; 2APS.
CC TIGR; NMB1398; --
CC InterPro: IPR001424; SOD_CU_ZN.
CC Pfam: PF00080; SOD_CU_ZN; 1.
CC PRODOM: P000469; SOD_CU_ZN; 1.
CC PROSITE; PS00087; SOD_CU_ZN; 1.
CC PROSITE; PS00332; SOD_CU_ZN; 1.
CC Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
KW SIGNAL 1 22 Potential.
FT CHAIN 23 186 Superoxide dismutase [Cu-Zn].
FT METAL 79 79 Copper (By similarity).
FT METAL 81 81 Copper (By similarity).
FT METAL 104 104 Copper and zinc (By similarity).
FT METAL 113 113 Zinc (By similarity).
FT METAL 122 122 Zinc (By similarity).
FT METAL 125 125 Zinc (By similarity).
FT METAL 160 160 Copper (By similarity).
FT METAL 186 182 By similarity.
FT DISULFID 86 182
FT CONFLICT 130 130
SQ SEQUENCE 186 AA; 19520 MW; 6499049BFA3427C CRC64;
Query Match 40.5%; Score 385; DB 1; Length 186;
Best Local Similarity 50.3%; Pred. No. 1.2e-26;
Matches 77; Conservative 22; Mismatches 52; Indels 2; Gaps 2;
Qy 28 SVTSEVIMDNGIKQSIGVTFTDCKGLQIKTDKGLPAGHGHFHHGGSCGPAEHD 87
Db 33 SIEVKVQQLDPVNGKDVGTFTTESNYGLVFTPDQLGSEGLHGHFHHENPCEPKE 92
Qy 88 GHLTAGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPRLT-VKEI 145
Db 93 GKLTAGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPRLT-VKEI 152
Qy 146 KGRVTMIHAGDNYSDKPLPLGGGARIACGVI 178
Db 153 RGHSTIMHTGDNHSDHPAPLPLGGGPRMACGVI 185
RESULT 37
SODC_PASMU STANDARD; PRT; 186 AA.
AC Q59689;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN Name=sodC; Ordered locus names=PM1952;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=PM70.
RC MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida PM70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
RL [2]
RN SEQUENCE OF 87-173 FROM N.A.
RP STRAIN=T5.
RC MEDLINE=96119708; PubMed=7496539;
RX Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RA "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RT from the eukaryotic enzyme, and not so rare after all!";
RL Microbiology 141:2271-2279 (1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB006232; AAK04036.1; --
CC EMBL; X83124; CAA58205.1; --
CC HSP; P24702; 2APS.
CC InterPro: IPR001424; SOD_CU_ZN.
CC Pfam: PF00080; SOD_CU_ZN; 1.
CC PRODOM: P000469; SOD_CU_ZN; 1.
CC PROSITE; PS00087; SOD_CU_ZN; 1.
CC PROSITE; PS00332; SOD_CU_ZN; 1.
CC Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
KW SIGNAL 1 20 Potential.
FT CHAIN 21 186 Superoxide dismutase [Cu-Zn].
FT METAL 79 79 Copper (By similarity).
FT METAL 81 81 Copper (By similarity).
FT METAL 104 104 Copper and zinc (By similarity).
FT METAL 113 113 Zinc (By similarity).
FT METAL 122 122 Zinc (By similarity).
FT METAL 125 125 Zinc (By similarity).
FT METAL 160 160 Copper (By similarity).
FT DISULFID 86 182
FT CONFLICT 130 130
FT CONFLICT 130 130 A -> V (in Ref. 2).
SQ SEQUENCE 186 AA; 19718 MW; 34D702070EE0DDC CRC64;
Query Match 40.2%; Score 382; DB 1; Length 186;
Best Local Similarity 51.7%; Pred. No. 2.3e-26;
Matches 77; Conservative 17; Mismatches 53; Indels 2; Gaps 2;
Qy 32 EVHIMDDNGIKQSIGVTFTDCKGLQIKTDKGLPAGHGHFHHGGSCGPAEHDGHLT 91
Db 37 KVELLDPAKGNQIGKGVITTESNYGLVFTPDQLGSEGLHGHFHHENPCEPKE 96
Qy 92 AGIQAAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPRLT-VKEI 149
Db 97 AGIQAAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPRLT-VKEI 156
Qy 150 VMTHAGDNYSDKPLPLGGGARIACGVI 178
Db 157 LMIHEGDNHSDHPAPLPLGGGPRMACGVI 185

```

RESULT 38
SODC ACTPL STANDARD; PRT; 190 AA.
AC P24702; Q59135;
DT 01-MAR-1997 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (BC 1.15.1.1).
GN Name=sodC;
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=715;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
RX MEDLINE=97101016; PubMed=8945543;
RA Langford P.R., Loyds B.M., Kroll J.S.;
RT "Cloning and molecular characterization of Cu,Zn superoxide dismutase
from Actinobacillus pleuropneumoniae.";
RL Infect. Immun. 64:5035-5041(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S 4074 / Serotype 1;
RA Helie M.C., Sirois M., Ouellet C., Boissinot M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 180-190 FROM N.A.
RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
RX MEDLINE=92158680; PubMed=1741300;
RA Loyds B.M., Langford P.R., Kroll J.S.;
RT "recF in Actinobacillus pleuropneumoniae.";
RL Nucleic Acids Res. 20:615-615(1992).
RN [4]
RP SEQUENCE OF 91-177 FROM N.A.
RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
RX MEDLINE=96118708; PubMed=7496539;
RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
from the eukaryotic enzyme, and not so rare after all!";
RL Microbiology 141:2271-2279(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=20124004; PubMed=10656823;
RA Forest K.T., Langford P.R., Kroll J.S., Getzoff E.D.;
RT "Cu,Zn superoxide dismutase structure from a microbial pathogen
establishes a class with a conserved dimer interface.";
RL J. Mol. Biol. 296:145-153(2000).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X99396; CAA67771.1; -
EMBL; U51440; AAB02816.1; -
DR EMBL; X63626; CAA45174.1; -
DR EMBL; X83123; CAA58204.1; -
DR F01; I39650; I39650.
DR DB; 2APS; X-ray; A/B=29-190.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; Sod_Cu_1.

```

```

DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW 3D-structure; Antioxidant; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
FT SIGNAL 1 23
FT CHAIN 24 190 Superoxide dismutase [Cu-Zn].
FT METAL 83 83 Copper.
FT METAL 85 85 Copper.
FT METAL 108 108 Copper and zinc.
FT METAL 117 117 Zinc.
FT METAL 126 126 Zinc.
FT METAL 129 129 Zinc.
FT METAL 164 164 Copper.
FT DISULFID 90 186
FT CONFLICT 113 113 E -> D (in Ref. 2).
FT CONFLICT 124 124 N -> D (in Ref. 2).
FT CONFLICT 134 136 FVE -> TIA (in Ref. 2).
FT STRAND 38 44
FT TURN 47 49
FT STRAND 52 62
FT TURN 63 64
FT STRAND 65 72
FT STRAND 78 80
FT STRAND 82 85
FT STRAND 93 95
FT TURN 96 97
FT STRAND 98 100
FT TURN 101 102
FT HELIX 103 105
FT STRAND 108 108
FT TURN 111 112
FT TURN 120 121
FT TURN 123 124
FT TURN 127 128
FT STRAND 129 129
FT TURN 133 135
FT TURN 137 138
FT STRAND 141 141
FT STRAND 145 147
FT TURN 148 149
FT HELIX 153 156
FT TURN 157 158
FT STRAND 159 164
FT HELIX 176 179
FT STRAND 183 190
SQ SEQUENCE 190 AA; 20207 MW; 5FC1F3148972DC83 CRC64;
Query Match 39.6%; Score 376; DB 1; Length 190;
Best Local Similarity 51.7%; Pred. No. 9e-26;
Matches 77; Conservative 18; Mismatches 52; Indels 2; Gaps 2;
QY 32 EVHMIDNGIKQSIGTGTFTDTKGLQIKTDKGLPAGEHGFIHGGSCGPAEHDCGLT 91
Db 41 QVQQLDPVKNKVDGTVETESAYGLVFTPLHGLAQGLHGFIHQNPSCPEKDKLV 100
QY 92 AGLQAHGHYDPDKTKEHGPIGL-NGHKGDLPLRVVVKADGIAKETLLAPRL-TVKEIKGT 149
Db 101 AGLGAGGHWPDKETQKHGYPMWDNAHLGDLPLFVHEDGSGATNPLAPRLKCLDEVKGS 160
QY 150 VMIHAGGDNYSDEKPLPGGGGARIACGVI 178
Db 161 LMIHEGDNHSDHAPLPGGGPRMACGVI 189
RESULT 39
Q6G3C6
ID Q6G3C6 PRELIMINARY; PRT; 175 AA.
AC Q6G3C6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Superoxide dismutase.

```

GN	Name=sodC; OrderedLocusNames=BH08570;	RA	Stevens M.K., Hassett D.J., Radolf J.D., Hansen E.J.;
OS	Bartonella henselae (Rochalimaea henselae).	RT	"Cloning and sequencing of the gene encoding the Cu,Zn-superoxide
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;	RL	dismutase of Haemophilus ducreyi.";
OC	Bartonellaceae; Bartonella.	RL	Gene 183:35-40(1996).
OX	NCBI_TaxID=38323;	[3]	
RN	[1]	RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 49882 / Houston 1;	RC	STRAIN=35000HP / ATCC 700724;
RC	PubMed=15210978;	RA	Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA	Alsmark U.C.M., Karlberg E.O., Legault B.-A., Ardell D.H.,	RA	Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RA	Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,	RT	"The complete genome sequence of Haemophilus ducreyi.";
RA	La Scola B., Holmberg M., Andersson S.G.E.;	RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RT	"The louse-borne human pathogen Bartonella quintana is a genomic	[4]	
RT	derivative of the zoonotic agent Bartonella henselae.";	RP	SEQUENCE OF 100-186 FROM N.A.
RL	Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).	RC	STRAIN=35000HP / ATCC 700724;
CC	-1- FUNCTION: Destroys radicals which are normally produced within the	RX	MEDLINE=96118708; PubMed=7496539;
CC	cells and which are toxic to biological systems (By similarity).	RA	Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
CC	-1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).	RT	"Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
CC	-1- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By	RL	from the eukaryotic enzyme, and not so rare after all!";
CC	similarity).	RL	Microbiology 141:2271-2279(1995).
CC	-1- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.	CC	-1- FUNCTION: Destroys radicals which are normally produced within the
CC	EMBL: BX897699; CAP27655.1; -	CC	cells and which are toxic to biological systems. May play a role
DR	InterPro: IPR001424; SOD_CU_ZN.	CC	in the interactive biology of organisms with their hosts and so
DR	Pfam: PF00080; SOD_Cu; 1.	CC	contribute to their capacity to cause disease.
DR	ProDom: PD000469; SOD_CU_ZN; 1.	CC	-1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
DR	PROSITE: PS00087; SOD_CU_ZN; 1.	CC	-1- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
DR	PROSITE: PS00332; SOD_CU_ZN; 1.	CC	similarity).
KW	Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.	CC	-1- SUBUNIT: Homodimer (By similarity).
SK	SEQUENCE 175 AA; 18861 MW; 5CCG3A4ADB35D8E CRC64;	CC	-1- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
Query Match 39.4%; Score 374; DB 2; Length 175;		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
Best Local Similarity 44.7%; Pred. No. 1.1e-25;		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
Matches 80; Conservative 27; Mismatches 62; Indels 10; Gaps 5;		CC	the European Bioinformatics Institute. There are no restrictions on its
QY	2 KIKLFFVTSTVSLTSTSVLACSVTSEVHMIDNGIKSIGVTFTDQKGLQIKT 61	CC	use by non-profit institutions as long as its content is in no way
Db	3 KILFLFAQVAFNCISN----VLADSTQVYKIELKNNKPIETIEENIYGLIFAP 58	CC	modified and this statement is not removed. Usage by and for commercial
QY	62 DLKGLPAGHGFIHEGSCGPAEHGHLTAGLQAHGHYDPDKTGHGGLG-NHKGDL 120	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
Db	59 NLSSLPEGFHGFHVNFPSCD--TKDG--VIGGAAGHYDPYQTKHGLGPNYINGHLGL 114	CC	-----
QY	121 PRLVVKADGTAKETLLAPRL-TVKEIKGRVTWTHAGDNDYDKPLPLGGCGARIAAGVI 178	DR	EMBL: X98737; CAA67289.1; -
Db	115 PALYVDEQGRATWSVAPRLKUSKSEVKGSHVIIHGDNQSDKPLPLGGGARLAGGII 173	DR	EMBL: U47664; AAB41293.1; -
RESULT 40		DR	EMBL: AE07153; AAP95739.1; -
SODC_HAEDU	STANDARD; PRT; 199 AA.	DR	EMBL: X83125; CAA58206.1; -
AC	Q59452; Q59449; Q59453;	DR	Pfam: PF00080; SOD_Cu; 1.
DT	15-DEC-1998 (Rel. 37, Created)	DR	InterPro: IPR001424; SOD_CU_ZN.
DT	15-DEC-1998 (Rel. 37, Last sequence update)	DR	ProDom: PD000469; SOD_CU_ZN; 1.
DT	05-JUL-2004 (Rel. 44, Last annotation update)	DR	PROSITE: PS00087; SOD_CU_ZN; 1.
DE	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).	DR	PROSITE: PS00332; SOD_CU_ZN; 1.
GN	Name=sodC; OrderedLocusNames=HD0848;	KW	Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
OS	Haemophilus ducreyi.	KW	Periplasmic; Signal; Zinc.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;	FT	Potential.
OC	Pasteurellaceae; Haemophilus.	FT	SIGNAL 1 22
OX	NCBI_TaxID=730;	FT	CHAIN 23 159
RN	[1]	FT	Superoxide dismutase [Cu-Zn].
RP	SEQUENCE FROM N.A.	FT	Copper (By similarity).
RC	STRAIN=35000HP / ATCC 700724;	FT	Copper (By similarity).
RC	MEDLINE=97288949; PubMed=9143881;	FT	Copper and zinc (By similarity).
RA	Langford P.R., Kroll J.S.;	FT	Zinc (By similarity).
RT	"Distribution, cloning, characterisation and mutagenesis of sodC, the	FT	Zinc (By similarity).
RT	gene encoding copper/zinc superoxide dismutase, a potential	FT	Zinc (By similarity).
RT	determinant of virulence, in Haemophilus ducreyi.";	FT	Zinc (By similarity).
RL	FEMS Immunol. Med. Microbiol. 17:235-242(1997).	FT	Copper (By similarity).
RN	[2]	FT	Copper (By similarity).
RP	SEQUENCE FROM N.A.	FT	DISULFID 99 195
RC	STRAIN=35000HP / ATCC 700724;	FT	By similarity.
RC	MEDLINE=97149276; PubMed=8996084;	QY	33 VHMIDNGKQSIGVTFTDQKGLQIKTDKGLPAGHGFIHEGSCGPAEHGHLTA 92
EX		Db	51 VQQLDPQNGKVDVGTVEITESAYGLVFTPKLHDLAHLGHFIHEKPSCEPKDKQLVA 110
		QY	93 GLQAHGHYDPDKTGHGGLG-NHKGDLPRLVKADGTAKETLLAPRL-TVKEIKGRVT 150

Db 111 GLGAGCHNDPKQTKHGYEWSDDAHMGDLPALFVNHGGSATTPVLAPRLKKLAEVKGHSL 170
QY 151 MIHAGGDNYSDKPLPLGGGGARIACGVI 178
Db 171 MIHAGGDNHSDEHPAPLGGGGPRMACGVI 198

Search completed: October 26, 2004, 09:44:55
Job time : 198 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 00:05:50 ; Search time 92 seconds
(without alignments)

Title: US-10-009-823A-1

Perfect score:

Sequence: 1 MMGSLFIGATGMKTHSTGLG.....NSKSVTTADTMLQKALELKR 502

Scoring table: BLOSUM62

Xgapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length:

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```

MODE=frame+g2n_model DEV=xlh
-o/sgn2_1/USPTO pool/US10009823/runat_26102004_100130_1634/app_query.fasta_1.647
-DE=issued Patents NA -QMT=fastap -SUPTXT=xrni -MINMATCH=0 -1 -LOCPCT=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blasmus62 -TRANS=human40_cdi
-LIST=45 -DALIGN=200 -THR SCORE=ECT -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pct -NORM=expct -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10009823 -CGEN 1 54 @runat_26102004_100130_1634 -NCPCT=6
-NO MMAP -LARGELQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONG=6
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREATS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DEVELOP=6 -DEEXT=7

```

Database : Issued Patents NA:*

- ```
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTCUS11.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query |        | Length | PB                  | ID                | Description |
|------------|-------|-------|--------|--------|---------------------|-------------------|-------------|
|            |       | Match | %      |        |                     |                   |             |
| 1          | 528.5 | 20.4  | 1416   | 4      | US-09-252-991A-2955 | Sequence 2955, Ap |             |
| 2          | 528.5 | 20.4  | 1512   | 4      | US-09-252-991A-2728 | Sequence 2728, Ap |             |
| C 3        | 475.5 | 18.4  | 1239   | 4      | US-09-543-681A-4010 | Sequence 4010, Ap |             |
| 4          | 414.5 | 16.0  | 640681 | 4      | US-09-750-998-1     | Sequence 1, Appl  |             |
| 5          | 307   | 11.9  | 810    | 4      | US-09-252-991A-3144 | Sequence 3144, Ap |             |
| C 6        | 302   | 11.7  | 1545   | 4      | US-09-253-991A-2823 | Sequence 2823, Ap |             |
| 7          | 286   | 11.1  | 798    | 4      | US-09-543-681A-3938 | Sequence 3938, Ap |             |
| 8          | 281.5 | 10.9  | 1800   | 3      | US-08-483-857-1     | Sequence 1, Appl  |             |
| 9          | 277.5 | 10.7  | 1800   | 1      | US-08-436-748-1     | Sequence 1, Appl  |             |
| 10         | 237.5 | 9.2   | 789    | 3      | US-08-483-857-3     | Sequence 3, Appl  |             |
| 11         | 223.5 | 8.6   | 789    | 1      | US-08-436-748-4     | Sequence 4, Appl  |             |
| 12         | 193   | 7.5   | 5445   | 4      | US-09-689-065B-2    | Sequence 2, Appl  |             |

## ALIGNMENTS

## RESULT 1

US-09-252-991A-2955  
; Sequence 2955, Application US/09252991A

Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELAT

; TITLE OF INVENTION: AERUGI

FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1998-03-18  
 ; CURRENT APPLICATION NUMBER: US/09/252,99

;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: IIS 60

; PRIOR APPLICATION NUMBER: US 60/014,788  
 : PRIOR FILING DATE: 1998-02-18

;  
: PRIOR FILING DATE: 1998-02-10  
: PRIOR APPLICATION NUMBER: US

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

NUMBER OF SEQ

; SEQ ID NO 2955

; LENGTH: 1416

TYPE: DNA

|                        |          |
|------------------------|----------|
| Alignment Scores:      |          |
| Pred. NO.:             | 4.66e-47 |
| Score:                 | 528.50   |
| Length:                | 1416     |
| Matches:               | 151      |
| Conservative:          | 86       |
| Mismatches:            | 205      |
| Indels:                | 75       |
| DB:                    | 16       |
| Percent Similarity:    | 45.84%   |
| Best Local Similarity: | 29.21%   |
| Query Match:           | 20.44%   |
| DB:                    | 4        |
| Gaps:                  | 16       |

US-10-009-823A-1 (1-502) x US-09-252-991A-2955 (1-1416)

4 SerLeuPheTlLeGlvAlaThrGlyMetIvstHrHisserThrGlyLeuGlyThrValSer 23

db 31 AGTTTCAACATCGGCTTGAGCGGCATTCAGGCGGCTCTAGCGGCCTGAACGTCAACGGC 90

```

QY 24 AsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
DB 91 AACAAATCGCCCAACCGCGGACCTGAGCTTCAAGCAATCCCGCGGAGTTCGCCGAC 150
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
DB 151 GTCTACGCGCTCCGGTG---CTGGGTTCG-----GGCAGCAACCCGCGAGGC 195
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGlnProGly 83
DB 196 AGCGGGTGCTCTCTCGACGCTCTCGCAGATGTTCAAGCAGGGCAATCAGCTCGAC 255
QY 84 AsnSerValThrAspLeuAlaIleGlyGlySerGlyPheGlnValThrLeuGluAsp 103
DB 256 AACAGCGTGCTGGAGCTGGCCATCAACGCGCAACGCGTCTCTCGTCACCAACACGCG 315
QY 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
DB 316 GCGATCAGTACACCGCGCGCGCTACTTCAATCCGCAAGCAGGATTTTCATGTCGAC 375
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlu 143
DB 376 AACAAACGGCTACCGGCTCGAGGCTATGCGCGCGCAACGCGGAGTTGCGACACGCG 435
QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
DB 436 GTGGTTCACGACCTCAAGTTCGAGCGCGCCATCAGCGCGCGCAGGCGCCCTCGAGCATC 495
QY 163 LysThrSerThrAlaLeuAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
DB 496 CAGCAGTGTCAACCTCACTCGACGCTGAAG-----CCG 531
QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
DB 532 CCGACCGTGAGCGCGTTC-----GATCCG 555
QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnValAsn 222
DB 556 TCCGACGCGGTACTACACTCTCTCTCTGCGTGGCATCTATGACACCGAGGCAAC 615
QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
DB 616 TCCACACCATGAGCCAGTCTTTCATCAAGAACGAGCGCGGACCGAATGCGACCGCGCG 675
QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
DB 676 ATTCGGGAGAACACGTGACCATGAAGTGTATCGACGCGTCAATCCGCTCGATCCG 735
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 736 TCGAACCAAGACGCG-----ATGAGCTTCAACGTCACCTTC 771
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
DB 772 GACCCAGCGCGCAGATGACCTGGTTCGGGCG-----CCGACGCGGACGACACGCGG 825
QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
DB 826 CCGGGCTTCAGCATCGACGCGGACCAACGATGATCCAGTTCAGTCGCGC----- 876
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
DB 877 ACCGGCAATCCGCGCATCC-----GGCACCCTGGTGGATTCGCGCGCGCTCG 924
QY 325 AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyValaProAlaSerAlaAla 344
DB 925 GAC---GGCAAGACCCCGCGCAGCTACGCTGGAATGGCGCGGACCGGTCCGCGCGCG 981
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
DB 982 ATTCCTTCACATGCGCAAG-----ACCAACCCAGTAC 1014
QY 365 SerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384

```

```

DB 1015 TCACCGCGGTTCGCCAGGACCAACCCGATC-----CAGGACGCTACACACC 1062
QY 385 GlyAspSerValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
DB 1063 GGTCACTCGCAGCGCTCGAAATCGACGACACCGGGTGATCTTCGCCGCTACACCAAT 1122
QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaAaGPhThrSerGluAspGlyLeu 424
DB 1123 GGCATCTCAAGGTGCGGCGCAGGTGGTGGCGAACTTCGCCAACATCCAGGGCTTG 1182
QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
DB 1183 ACGCCGATCGGCAAGACCTCTCTGGGTGCGATCCGCGGAGTCCGCGGAGCGCGGTGCG 1242
QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
DB 1243 GCGCGCGCTCGCGGCAACCTCGGGGCGGTTCGAATCCGCGCGCTGGAAGCGTCCAACGTG 1302
QY 465 AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484
DB 1303 GACATCTCCAACGAACTGGTGAACCTCATCGTCCACGAGCGCACTACCGGCAATGCC 1362
QY 485 LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
DB 1363 AAGACCATCCAGACGAGATGCGGTGACCCAGACCATCATCAACCTGCGC 1413

RESULT 2
US-09-252-991A-2728/c
; Sequence 2728, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2728
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2728

Alignment Scores:
Pred. No.: 5,19e-47 Length: 1512
Score: 528.50 Matches: 151
Percent Similarity: 45.84% Conservative: 86
Best Local Similarity: 29.21% Mismatches: 205
Query Match: 20.44% Indels: 75
DB: 4 Gaps: 16

US-10-009-823A-1 (1-502) x US-09-252-991A-2728 (1-1512)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
DB 1503 AGTTTCAACATCGCCCTGAGCGGCATCCAGCGCGCTCTAGCGGCTTGAACGTCACCGGC 1444
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
DB 1443 AACAAATCGCCCAACCGCGGACCTGAGCTTCAAGCAATCCCGCGGAGTTCGCCGAC 1384
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
DB 1383 GTCTACGCGCTCCGGTG---CTGGGTTCG-----GGCAGCAACCCGCGAGGC 1339
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGlnProGly 83

```

Db 1338 AGCGGGTCTCTCTCGACGCTCTCGCAGATGTTCAAGACGAGGCAACATCGACTCGACC 1279  
QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluasp 103  
Db 1278 AACACGGTGTGACCTGCGCATCAACGGCAACGGCTTCTTCCTCACCAGCAACACGGG 1219  
QY 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123  
Db 1218 GCGATCAGTACACCGCGCGCGCTACTCAATACCGACAGCAGGATTCATCGTCGAC 1159  
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143  
Db 1158 AACACAGGCTACCGCTGTCAGGGCTATCGCTCGCGCGCAACGGCCAGTTCAGACAGCGC 1099  
QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162  
Db 1098 GTGGTCACCGACCTCAAGTCCAGCGCGCCCAATCAGCGCGCGCAGGCACCTCGAGCATC 1039  
QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182  
Db 1038 CAGCAGTCTGTACAACCTCACTCGACGCTGAAG-----CCG 1003  
QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202  
Db 1002 CCGACCGTGACCGCTTC-----GATCCG 979  
QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222  
Db 978 TCCGACGCGCTTACCTACCACTCGCTCTTCCTCGCGCATCTATGACAGCGGCAAC 919  
QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237  
Db 918 TCCACACATGACGACGCTTCTCATCAAGAACGCGCGGACCGCAATGGACCGCGCG 859  
QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253  
Db 858 ATTCCGGAGAACAGCTGACCATGAAAGTGTGATCAGCGCGCTCAATCGCTCGATCCG 799  
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuMetSerGlyThrMetThrPhe 273  
Db 798 TCGAACAGACGCGC-----ATGAGCTTCAACGTCACCTTC 763  
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290  
Db 762 GACGCCAGCGGCAGATCGCTCGTTCGGGCG-----CCGACGCGCAGCAGCGGG 709  
QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304  
Db 708 CCGGGCTTCAGATCGACGCGCAGCACCACCACTGATCCAGTTCAGTCCGGCC----- 658  
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324  
Db 657 ACCGGCAATCCGCGACTCCC-----GGCACCCTGGATTCGGCGGCTCG 610  
QY 325 AppPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344  
Db 609 GAC---GGCAGACACCCCGCACCTACCGCTGGAATCGCGCAGCCGCTCGCGCAGCGGC 553  
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364  
Db 552 ATCTCTTCGACATCGCAAG-----ACCACCCAGTAC 520  
QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 394  
Db 519 TCCACCGGTTCGCCAGACGACCCCGATC-----CAGGACGGCTACACCATC 472  
QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404  
Db 471 GGTACGTGGCAGCGCTGGAATCGACGACACCGCGGTGATCTTCGCCCGCTACACCAT 412  
QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424  
Db 411 GCCCATCTCAAGGTGACGGCGCAGGTGTGCTGCGCAACTCGCCACATCCAGCGGCGCTG 352

QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444  
Db 351 ACGCCGATCGGCAAGACCTCTCTGGTGCAGTCCGCGAGTCCGCGAGCGGGCGGCG 292  
QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464  
Db 291 CGCGCGCTCGGACCCCTCGGGCGTTGCAATCCGCGCGCTGGAGAGCTCAACGTG 232  
QY 465 AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484  
Db 231 GACATCTCAAGCAACTGGTGAACCTCATCGTCCACCAGCGCACTACCAAGGCAATGCC 172  
QY 485 LysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501  
Db 171 AAGACCATCCAGACGAGATGCGGTGACCCAGACCATCATCACTGCGC 121  
RESULT 3  
US-09-543-681A-4010  
; Sequence 4010, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543.681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4010  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4010  
Alignment Scores:  
Pred. No.: 1,958-41 Length: 1239  
Score: 475.50 Matches: 136  
Percent Similarity: 42.63% Conservative: 81  
Best Local Similarity: 26.72% Mismatches: 165  
Query Match: 18.39% Indels: 127  
DB: 4 Gaps: 15  
US-10-009-823A-1 (1-502) x US-09-543-681A-4010 (1-1239)  
QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29  
Db 40 AGTGGTTAAATGACGAGCGCGCTAAATTAGATACTATTGGTAATAATATCTCAAACTCT 99  
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
Db 100 GCACTTACGTTTAAAGGGGCAACAGTCTCTTTCCGATGTTTTCG----- 150  
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69  
Db 151 -----GGCTCT-----GGTGCAGGGCTGTGTGTAAAGATATCA 183  
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89  
Db 184 GGCATTAGCCAAACTTTAAAGATGGTAGTATCACTACTACTACCGCCCAACCGATGTC 243  
QY 90 AlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106  
Db 244 CGCATTTCTGTGTGTGTTCTTCCTCGTATT-----GAAGATAGTAAACGCGTGGCGTTTC 297  
QY 107 TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126  
Db 298 TATTACGTAACGCTGAATTGGCAAGATAAAGCGGTATCTCTCAATAATGCAAGGC 357  
QY 127 PheThrLeuMetGlySerArgIleSerAsn-----AsnProhenIleLysLys 142

Db 358 ATGCGTATTACAGTTATCCAGTACAAACAGGTAGACGTTAAATAATGTAGTACAAAAGGG 417  
QY 143 GluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAla 162  
Db 418 GCAACACCAACACCTCATATT-----CCTACCGATATGATGAAT---GCA 462  
QY 163 LysThrSerThraLeuAsnAlaValValAsnLeuGly-----AspSerThrAsp 179  
Db 463 AGTCGAACCGATAAATATGATATGATGCGTTAACTTGAATTCAGCTGACAGCGCTATCGAT 522  
QY 180 LysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyrLysGlyAsnGly 199  
Db 523 CAACACACATAAATTGATCCC-----AAGATATAC--- 555  
QY 200 ThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGln 219  
Db 556 -----GATTCCTATACTTACTACTACGTACGTACCAACCTACGATAGC 597  
QY 220 GlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySer 239  
Db 598 TTAGTAAATGACATACCTGACTTATCTTT-----GTTAAACCC 639  
QY 240 LysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly 259  
Db 640 AAAGATAACGAATGAGTGTATGCTCAAGACACCAACAGGTGAGCCAGCACAGAT 699  
QY 260 ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGly----- 277  
Db 700 CTG-----GGTAAATTAGTTTATAGGATAACGGGGTATTA 735  
QY 278 -----GluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292  
Db 736 GATGAACACGACCTAAGCTGAAACCTTTACTACCGTGTCTTATAAGGTTCA----- 789  
QY 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312  
Db 789 ----- 789  
QY 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332  
Db 790 -----CARACCATGGATATGGAATGACTTT----- 816  
QY 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352  
Db 816 ----- 816  
QY 353 ProSerMetProIleGlnThrSerSerGlyAsnSerThrAlaAlaAsnGlySerSer 372  
Db 817 -----AGTGGTAGTACGACGACCAAAAGTCGCTGATCA 849  
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392  
Db 850 AGTGTCTCTAAACTCGCGCAAAATGTTATCAAGCAGGGGAATTTACCAATTTCCGTATT 909  
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412  
Db 910 GAGCCAGATGTTCCATATATGCGACCTACTCAACCAACCAACCAAGCAAGTGGTGGTCAA 969  
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432  
Db 970 ATCGCATTAGCTAATTTGCAATCTCGTGGCTTAAGTTCAAGGTGACATATGTTGG 1029  
QY 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452  
Db 1030 TCTGAACCAATGGTTCAGGCTCACCATTGTTGGTGTGCGAGGCTCTGGCGTATTCGT 1089  
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472  
Db 1090 AAATTAACCAACAATGCTTTAGAGCTCTACGTAGATATGAGCCAGAGGTAGTTAAC 1149  
QY 473 MetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492  
Db 1150 ATGATTGTTGCTCAACGTAACATCAATCTAATGCGCAGACTCAAAACACTCAGGATCAG 1209

QY 493 MetLeuGlnLysAlaLeuGluLeuLys 501  
Db 1210 ATCTTGCACACACTAGTTAGCATGCGC 1236

## RESULT 4

US-09-790-988-1  
Sequence 1, Application US/09790988  
Patent No. 6632935  
GENERAL INFORMATION:  
APPLICANT: SHIGENOBU, SHUJI  
APPLICANT: MATANABE, HIDEKI  
APPLICANT: HATTORI, MASAHIRO  
APPLICANT: SAKAKI, YOSHIYUKI  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09790,988  
CURRENT FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: JP2000-107160  
PRIOR FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Alignment Scores:  
Pred. No.: 2,038-30 Length: 640681  
Score: 414.50 Matches: 132  
Percent Similarity: 39.96% Conservative: 73  
Best Local Similarity: 25.73% Mismatches: 187  
Query Match: 16.03% Indels: 121  
DB: 4 Gaps: 13

US-10-009-823A-1 (1-502) x US-09-790-988-1 (1-640681)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23  
Db 372153 TCAATATGATAGCTATTAAGTGGCTTACTAGCAATAATATGATTACATGGAAATATATATCC 372212  
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValAlaPheGlnAsp 43  
Db 372213 AATAATATGCTAACGCATCAACTATAGGATATAAATCTCGTAAACCTCTTTTGTGAT 372272  
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63  
Db 372273 ATGTTTCTCATTCATTTTATTCAAATACTACTAAT-----GGA 372311  
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83  
Db 372312 TACGGAGTGGGTATTTCAAGTATTTATACAAACTTTTAAATGCGCATGTTAGTTGAAACT 372371  
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGln---ValThrLeuGlu 102  
Db 372372 CGACGAGATTGGATTTAGGAATTTATAAAGACGCGCTTTTTCGCTCTTTAGACAGTCAA 372431  
QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122  
Db 372432 GGTCTATGCTCTATATACAGAGATGGCAATTTCTTCCTCATTAAGATCAAAATATTATC 372491  
QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136  
Db 372492 AATATTCAGGTATGATCTTAACCTGGACTTATACATCTTGTTCAAAAGCTGATTTTAAAT 372551  
QY 137 AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156  
Db 372552 AAT-----AGATCCAACTTAGAACCTTATTAAATTA-----AAAAATTCCTAAT 372593  
QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly--- 175  
Db 372594 ATTTTAAAAAACCAACCACTACTCTCTGAAATCATCTTAAAGCGCTTTTGTGAATCGTAATACT 372653

QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189  
Db 372654 GAATCAAAAAGCAGTGTGATTAATCTGACAACTATCTAACACGAGATTACATG 372713  
QY 190 AlaLeuLeuGluSerTyrPheGlyAsnGlyThrProProlleSerThrSerAsnTyrSer 209  
Db 372714 ACTTATATAGC----- 372725  
QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229  
Db 372726 -----ATATATAATAAGAGGGAAGGAAAAAGAGATATATCTGTTCT 372767  
QY 230 PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn 249  
Db 372768 TTTAAT-----AAAAAGAAACAATAAATGGAACAGTAAAT-----GTGGAATCAAAAT 372815  
QY 250 ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly 269  
Db 372816 GATTCTGATGATAAGAGACTATAAATAAT-----AGTTTC 372851  
QY 270 ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly 289  
Db 372852 GATTTAACGTTAATGATGATGGCAATTAACCTCTGATTAATGTTTAAATATATACATCT 372911  
QY 290 SerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGln 309  
Db 372912 AAAGATTCTAAAAG----- 372926  
QY 310 PheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyIleLys 329  
Db 372927 -----TATGAAATATCACTTTAAATTTA----- 372950  
QY 330 SerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349  
Db 372950 ----- 372950  
QY 350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369  
Db 372951 -----ACAGTACTATAGAACCAATCAAT 372974  
QY 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389  
Db 372975 TCTGATGTTTCTGGGAAGAACATCTCAAAACGATACCTCAAGTAAATTTAAACA 373034  
QY 390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409  
Db 373035 TTTGATATTGTTACTAATGTTGGAATATTGGAACATATTGCAATCAAAACCAACA 373094  
QY 410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsn 429  
Db 373095 ATAGTCAATATTATTATCAAAATTTATCAATCCAGAAATTTACAACTGAAAGTGGT 373154  
QY 430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu---ProGlyThr 448  
Db 373155 AATTATGCTGCTGCTACTGCAAGATCAGTGAAGCAAAACAGCAATGAAAGCGGTATT 373214  
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArg 468  
Db 373215 CAAGATCAGAGTGTAGCAATTAAGCAATAAAGCGTAGAAGTCAAAATGTTGATTTGAATAAA 373274  
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488  
Db 373275 GAAATTAATCAATATGATTATAGCAACAGTAATATCAATCAATCAATCAATCTTTTAA 373334  
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501  
Db 373335 ACAGAGATAAATAATTAATCAATTAATAATTTACAG 373373

## RESULT 5

US-09-252-991A-3144

; Sequence 3144, Application US/09252991A

; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3144  
; LENGTH: 810  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3144

## Alignment Scores:

Pred. No.: 1,456-23 Length: 810  
Score: 307.00 Matches: 108  
Percent Similarity: 30.74% Conservative: 46  
Best Local Similarity: 21.56% Mismatches: 93  
Query Match: 11.87% Indels: 254  
DB: 4 Gaps: 11

US-10-009-823A-1 (1-502) x US-09-252-991A-3144 (1-810)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20  
Db 25 ATGTATCGGCACCTGTGGTTCAGCAAGACCGGTCTGTCCGCCAGGACATGAACCTGACC 84  
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40  
Db 85 ACCATTCCCAACACTGGCCACGATATCCACACCGGCTTCAAGCGCGACCGCGGAG 144  
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60  
Db 145 TTCAGGACCTGTGTATCCAGATCCGCGCCAGCGCGCGGCGGACGCCAGGACAGC 204  
QY 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79  
Db 205 GAGTGCCTTCGGGCTGCACTGACCTGGGTACCGGTGTGGCGTGTGCGGACCCAGAGATC 264  
QY 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94  
Db 265 TTCACCGCGGCGCAGCTGTGACAGCACCGAGCAGCCGCTGGACATGGCGGTCAACGGCGC 324  
QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113  
Db 325 GCTTCTTCCAGGTCTGCTGCGGACGGACCGCTGTCTACACCGCGCGGACGCTTC 384  
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133  
Db 385 CACCTGAACCTCCGACGGCGGACATCGTCACCTCCACGGCTTCGCCCTG----- 432  
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153  
Db 432 ----- 432  
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173  
Db 433 GAGCCAGCGATC----- 444  
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193  
Db 444 ----- 444  
QY 194 SerTyrLysGlyAsnGlyThrProProlleSerThrSerAsnTyrSerTyrAlaGlnPro 213  
Db 444 ----- 444  
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233

```

Db 444 ----- 444
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 445 ----- 445
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 460 ----- 460
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 472 ACC----- 474
QY 294 AspLeuAlaAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 474 ----- 474
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 475 ---GTCGGC----- 480
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 480 ----- 480
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
Db 480 ----- 480
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 481 ----- 481
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 505 ACCACCGGCAACGCCAG-----CCGAGGTGATCGGC-----AACATC 543
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 544 CAGACCGCGGATTCATCAACCCCGCGGCTGCGAGGCCATCGCAACACCTGTTCTCTG 603
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 604 GAACCGGCTCCAGCGCGCGCTACGAGTGGTACGCGGCTCTCAACCGGCTCGGCACG 663
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 664 GTTCCCAAGAACACCTCGGAACCTCAACGTCACGTCGAGGAACTGTTGAACATG 723
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 724 ATACCAACCGAGCGCGCTACGAGATGAACTCAAGTCAATCTCCACCGCGCACGATG 783
QY 494 Leu 494
Db 784 TTG 786

```

## RESULT 6

```

US-09-252-991A-2823/c
; Sequence 2823, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2823
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2823

```

```

Alignment Scores:
Pred. No.: 1.44e-22 Length: 1545
Score: 302.00 Matches: 107
Percent Similarity: 30.60% Conservative: 46
Best Local Similarity: 21.40% Mismatches: 93
Query Match: 11.68% Indels: 254
DB: 4 Gaps: 11

US-10-009-823A-1 (1-502) x US-09-252-991A-2823 (1-1545)

QY 2 MetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThr 21
Db 1545 TTATCGGCACTGTGGTTCAGCAAGACCGGTCTGTCCGCCAGGACATGAACCTGACACC 1486
QY 22 ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe 41
Db 1485 ATTTCCAAACACTGCGCAACGTATCCACACCGGCTTCAAGCGCGGAGGATTC 1426
QY 42 GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln 61
Db 1425 CAGGACCTGCTGTACCATCGCGGCCAGCGCGGCCAGTCGACCCAGGACGCGAG 1366
QY 62 AlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAlaPhe 80
Db 1365 CTGCTCTCGGCGCTCAACTGGTACCGGTGTCGCGTGTGCGCGACCCAGAGATCTTC 1306
QY 81 GluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLysGly 95
Db 1305 ACCCGCGGCGCTGTCAGACACCGAGCGCGCTGGACATGGCGGTCAACGGCGCGGCG 1246
QY 96 PhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 1245 TTCTTCCAGGTCTGCTCGCGGACCGGACCGTGTCTTACACCGCGCGGAGCTTCCAC 1186
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 1185 CTGAACCTCCGACGCGGAGATGCTCACCTCCAAACGGCTTCGCCCTG----- 1141
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 1140 -----GAG 1138
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 1137 CCAGCGATC----- 1129
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db 1129 ----- 1129
QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 1129 ----- 1129
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
Db 1129 ----- 1129
QY 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
Db 1128 -----GTGGTGGCCCAACGAG----- 1114
QY 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
Db 1113 -----ACCCAGACCTCTCACC 1099

```

```
QY 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 1099 -----
QY 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 1099 -----
QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 1098 GTCGGC-----
QY 335 TrpAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 1093 -----
QY 355 MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr 374
Db 1093 -----
QY 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
Db 1092 -----CAGGACGGC-----
QY 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
Db 1065 ACCGGCAACCCACG-----CCGAGGTGATCGC-----AACATCCAG 1027
QY 415 LeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAla 434
Db 1026 ACCGCCACTTCATCAACCCGGCGCGCTCGAGGCCATCGCACCAACCTGTTCTCTGGAA 967
QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
Db 966 ACCGGCTCCAGCGCGCGCGCGAGGTGCGTACCGCGGTCTCAACGGCCTCGGACGGTT 907
QY 455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
Db 906 GCCCAGAACACCTTGGAAACTCCACGTCAGCTGAGGAGTCTCGAGGAACTGGTGAACATGATC 847
QY 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
Db 846 ACCACCCAGCGCGCTACGAGATGAACCTCAAGTCTATCCACCGCCGACGAGATGTTG 787
```

## RESULT 7

```
US-09-543-681A-3938
; Sequence 3938, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3938
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3938
```

```
Alignment Scores:
Pred. No.: 2,6e-21 Length: 798
Score: 286.00 Matches: 100
Percent Similarity: 30.57% Conservative: 55
Best Local Similarity: 19.72% Mismatches: 98
Query Match: 11.06% Indels: 254
DB: 4 Gaps: 9
```

```
US-10-009-823A-1 (1-502) x US-09-543-681A-3938 (1-798)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 16 ATGATCCGCTCTTATGATTGCTAAACACAGGGTGGATGCACCAACAACTAAATGAT 75
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 76 GTGATTCCACCAACCTCGCAACGTCAGCACAAATGGTTTTTAACGCCACGCGGTT 135
QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGlySer 56
Db 136 TTTGAGGATTTACTCTATCAAACTATTGTCGCAACGGGAGCGATGACATCCGACGACG 195
QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 196 AATGGCGCTTCTGGTTTCAAAATGGTACTGCTGTTGCCCCAGTGGCAGACAACTTTA 255
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
Db 256 CATAGCCCAAGTAATTTAGCCCAAACTAATGGTACCGGTGATGTTCTATCAAGGGCAA 315
QY 95 GlyPhePheGlnValThrLeuGluAspLysValHis--TyrThrArgAlaGlyAsnPhe 113
Db 316 GGTITTTTCCATGTTCAATTTACCTGATGGTACGATGCTTATCTCGTATGGTCTTTT 375
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 376 CAAATGGACCAAAATGGCAACTAGTGACCTCCAGTGGCTTTCAAATCGTCCACGCG-- 432
QY 134 IleSerAsnAsnProAsnIleLysGlnThrLeuGluProIleGlnLeuPheAsn 153
Db 433 -----ATTATTTTGCAGAAACCGCTAAAGAGGTGATGGTA----- 468
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 468 -----
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
Db 468 -----
QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 468 -----
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 469 -----GGTCGTGATGGTATTGTCAAGTGTGGAGATTGAGAGATCG 507
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 508 CTGCA-----
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 513 -----
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 513 -----
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 514 -----CCTCAA-----
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 519 -----
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 520 -----CAAGTAGGGCAATTA--- 534
```

QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373  
Db 534 ----- 534  
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393  
Db 535 -----ACCTTCACC 543  
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413  
Db 544 ACA----- 546  
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433  
Db 547 -----TTTATTATGATAGCGGGTTAGAAAGTGTGGGAAATCTGTACTTA 594  
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453  
Db 595 GAAACAGCCAGCTCCGAGACCCCACTGAGAAATCGCGGGTATTAAACGGTGGCGGCTTG 654  
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473  
Db 655 TTATATCAGGATATGTTGAACCTCTAACGTTAATGTCGCGGAAGAAATGGTCAATATG 714  
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493  
Db 715 ATCCAACTCAGCGTCTTATGAATTAATAGTAAAGGATTCACACTTCTGATCAGATG 774  
QY 494 LeuGlnLysAlaLeuGluLeu 500  
Db 775 TTACAGAACTCAGCAACTC 795

## RESULT 8

US-08-483-857-1  
; Sequence 1, Application US/08483857  
; Patent No. 6020125  
; GENERAL INFORMATION:  
; APPLICANT: Chan, Voon Loong  
; APPLICANT: Louie, Helena  
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF  
; TITLE OF INVENTION: CAMPYLOBACTER  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/483,857  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-504  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1800 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:

; NAME/KEY: CDS  
; LOCATION: Join(101..910, 942..1730)  
; US-08-483-857-1

Alignment Scores:  
Pred. No.: 3 01e-20 Length: 1800  
Score: 281.50 Matches: 133  
Percent Similarity: 38.41% Conservative: 94  
Best Local Similarity: 22.50% Mismatches: 228  
Query Match: 10.89% Indels: 136  
DB: 3 Gaps: 22

US-10-009-823A-1 (1-502) x US-08-483-857-1 (1-1800)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20  
Db 101 ATGCAAAATGGATATTATCAAGCAACTGCGGAATGGTAACCTCAGTTTAAATAAACTTGTAT 160  
QY 21 ThrValSerAsnAsnIleAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40  
Db 161 GTGATTCTAATAATCTTGGCAATATCAATCACTGATATATAAGAGATGATGTGTT 220  
QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53  
Db 221 ATTGAGATTTTAAAGAGATTTTAAAGAACTCAGGATGAGTTGCCCTATAGAAAATCAC 280  
QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70  
Db 281 ACAAGAGATGATCTCTGTTTGTAAATACTCAATAGATGGAATCCCAAGTTTCTCAA 340  
QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90  
Db 341 GAAATACGAGATTTAGCTAGTTCTTAAAGGCGCACAAACAATCCTTTCGATTGGCA 400  
QY 91 IleGlyGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysValHisThr 108  
Db 401 ATGACTAGAGAGATGCTTTTATTTGTTGTCAGACCAAGATGAGAGATGAAGATTAAAC 460  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 461 AAAGATGGAATTTTCAACTTGATGATGAGGTTATTGGTAAATAAGCAAGGATACAAG 520  
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsn----- 139  
Db 521 GTATTAAAGTAGTATGATTATTTTAAATATCTCAGAAATGCTGCATACCATTCCTAATAGT 580  
QY 139 ----- 139  
Db 581 GCTGTTCAATTAGCGTTGATAAAAACGGAAGCATTGAAAGTTGATGGAGTCAAAATGCA 640  
QY 140 -----IleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn----- 153  
Db 641 AGATTATTGTCAGCACAGATGATATAGAGCTTTGCAAAAAGATGGGGATAATGTC 700  
QY 154 -----AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170  
Db 701 TATAAATAGATGATCTAACCCGTTATTAGAGATTGAAATAAATACTCAATGCTATTTCGCCAA 760  
QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSer-----GluAla 185  
Db 761 GGTTTTCTCAGGGATCAATGTTAATCCAGTTACTGAAATGGTAGGACTGATTGAAGCA 820  
QY 186 AsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThr 205  
Db 821 AAC-----AGAAATGGTAGAATGTATCAAAAAGTTATGACAGCTCATATGGATGAC 871  
QY 206 SerAsnTyrSerTyrAlaGlnProMet----- 214  
Db 872 TTAATCAAGAAGCTATCAATAAGCTTGCAGCTGTTAAATAATTTAAATAATAATAAAAA 931  
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234  
Db 932 AGGATTAATAAATGATGAGATCACTTCATAC-----TGCTGCTACAGGAATGGTAGCGCA 985



[illegible]

```

422 pGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProG1 442
1490 GGGTCTTCATTCTATCGGTGATAATCTTTATCTTGAACAGGAGCAAGTGGTGACCTGT 1549
442 uPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSe 462
1550 TCGGGGTATAGCAGGACAAGATGGGCTTGGACAATAAGACATGGATTTATAGACTTAG 1609
462 rAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMe 482
1610 TAAATGTTCACTTGTTCGAAGAATGACAGATCTTATCACAGGACAAAGAGCTTATGAAGC 1669
482 tAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysAr 502
1670 GGGTTCATAGGCCATACCAACAGATGATGATATGCTAGGAATTTGTAATCAGCTTAAGCG 1729
502 g 502
1730 A 1730

RESULT 10
US-08-483-857-3
; Sequence 3, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-483-857-3

```

Db 1 ATGATGAGATCACTTCTATCTGCTGCTAGCAAGTGGTAGCGCAGCAAAACAAATTCAT 60  
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrIysGlnGlnValVal 40  
Db 61 GTTACTTCAATCAATCAATCGCAATGTTAATACAGCAGGTTTAAAGAAAGTCCGCGAGAA 120  
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60  
Db 121 TTTGCTGATCTTATGATCAAGTTATGAGTATGACAGCACTTCAACTTCAGCTACTACT 180  
QY 61 GlnAlaGlyMetGlyAlaGlnValGly-----SerValargThrIle 74  
Db 181 CTTTCTCCTTCGGGTATAGAGTGGGTGGGTGGTGGTCCACAGCGGTAACTAAAGTT 240  
QY 75 PheThrGlnGlyAlaPheGlu-----ProGlyAsnSerValThrAspLeuAlaIleGlyGly 93  
Db 241 TTTACTGAAGAAATTTAAATCAACACAGTACTGATGGTCTTATGATGCTATTCAGCT 300  
QY 94 LysGlyPheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsn 112  
Db 301 AATGGGTGTTTTTCAATACAACTTCCTGATGGCACTATAGGATATAGTAATGGGCAA 360  
QY 113 PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySer 132  
Db 361 TTTACAAAGATAATGAGTAATATTGTAATTCAGATGCTTATAGACTT----- 411  
QY 133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152  
Db 411 ----- 411  
QY 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172  
Db 411 ----- 411  
QY 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192  
Db 411 ----- 411  
QY 193 GluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGln 212  
Db 411 ----- 411  
QY 213 ProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGly 232  
Db 411 ----- 411  
QY 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252  
Db 411 ----- 411  
QY 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThr 272  
Db 411 ----- 411  
QY 273 PheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292  
Db 411 ----- 411  
QY 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312  
Db 412 ----- 420  
QY 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332  
Db 421 ----- 438  
QY 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352  
Db 439 -----GCAACAGCAATTAATGTTGTCTACAGAT---GGAAACCGTT 474  
QY 353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372

Db 475 TCTGTAAATGCTACCA----- 489  
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392  
Db 489 ----- 489  
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412  
Db 490 -----GGGAGCAACAAGAACTCAAT-----GGCCAA 519  
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432  
Db 520 GTGGAGCTAGTTCAAGTTTATAATCCAGCGGTCTTCTATCTATGGGTGATAATCTTTAT 579  
QY 433 SerAlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452  
Db 580 CTTGAAACAGGAGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 639  
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472  
Db 640 ACAATAAGACATGGATTTATAGAACTTAGTAATGTTCAAGCTTGTGAAGAAATGACAGAT 699  
QY 473 MetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492  
Db 700 CTTATCACAGGACAAAGAGCTTATGACGCGGTCTTAAGCGGCTTACAGCAAGTGTAT 759  
QY 493 MetGlnGlnLysAlaLeuGluLeuLysArg 502  
Db 760 ATGCTAGGAATGTAAATCAGCTTAAGCGA 789

RESULT 11  
US-08-436-748-4  
; Sequence 4, Application US/08436748  
; Patent No. 5827654  
; GENERAL INFORMATION:  
; APPLICANT: CHAN, YOON LOONG  
; APPLICANT: LOUIE, HELENA  
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF  
; TITLE OF INVENTION: CAMPYLOBACTER  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/436,748  
; FILING DATE: 05-AUG-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, MICHAEL I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-428 MIS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 789 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-436-748-4

Alignment Scores: 1.4e-14 Length: 789  
Pred. No.:

Score: 223.50 Matches: 98  
 Percent Similarity: 27.70% Conservative: 43  
 Best Local Similarity: 19.23% Mismatches: 115  
 Query Match: 8.64% Indels: 253  
 DB: 1 Gaps: 12

US-10-009-823A-1 (1-502) x US-08-436-748-4 (1-789)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20  
 DB 1 ATGATGAGTCACTTCATCTCTGCTACAGAAATGGTAGCGCAGCAACAAATGAT 60  
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40  
 DB 61 GTTACTTCAATAACATCGCAATGTTAATACAGCAGGTTTTAAGAAAGTCGCGCAGAA 120  
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60  
 DB 121 TTGTCTGATCTATGATATCAAGTTATGAAGTATGATGAGCAACTTCAACTTCAGCTACTACT 180  
 QY 61 GlnAlaGlyMetGlyAlaGlnValGly-----SerValArgThrIle 74  
 DB 181 CTTTCTCCTCGGTATAGAGTGGTGTGGTGTGCTGCCAACAGCGGTAACTAAAGTT 240  
 QY 75 PheThrGlnGlyAlaPheGlu---ProGlyAsnSerValThrAspLeuAlaIleGlyGly 93  
 DB 241 TTTACTGAAGGAATTTAAATCAACAAAGTACTGATGCTTGTATATGGCTTATTCGAGGT 300  
 QY 94 LysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPhe 113  
 DB 301 AATGGGTTTTTCAATCAACTT----- 324  
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133  
 DB 325 -----CCTGATGCGCACTATAGAAATGGGC----- 348  
 QY 134 IleSerAsnAsnProAsnIleLysGluThrLeuGluProIleGlnLeuAspPheAsn 153  
 DB 349 -----AATTTACAAG----- 360  
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173  
 DB 361 -----ATAATCAAG----- 369  
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193  
 DB 369 ----- 369  
 QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213  
 DB 369 ----- 369  
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233  
 DB 370 -----GATAATCAGGTAATATT----- 387  
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253  
 DB 387 ----- 387  
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273  
 DB 388 -----GTAATTCAGATGGTTATAGACTTTTACCTGAATGACAATA 429  
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293  
 DB 430 -----CCTGAAGGC----- 438  
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313  
 DB 438 ----- 438  
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333

DB 438 ----- 438  
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353  
 DB 439 -----GCAACAGCAATTAATGTGTACAGAT---GGAACCGTTTCT 477  
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373  
 DB 478 GTRATGCTACCA----- 489  
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393  
 DB 489 ----- 489  
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413  
 DB 490 -----GGGAGCAACAAAGAACTCAATT-----GCCCAAGTG 522  
 QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgAggGluGlyAsnAsnHisTyrSer 433  
 DB 523 GAGCTAGTTTCAGTTTATAATCCAGCGGTCTTCATTCTATGGGTGATAATCTTTATCTT 582  
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453  
 DB 583 GAAACAGGAGCAAGTGGTGCACCTGTTGGGTATATGACGAGCAAGATGGGCTTGGAAACA 642  
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473  
 DB 643 ATAAGACATGGATTATAGACTTATGTTTTCAGCTTGTGAGAATGCACAGATCTT 702  
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493  
 DB 703 ATCAGACAGCAAGAGCTTATGACGGGTCTTACAGCGCATTAACAACAGTATGATATG 762  
 QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502  
 DB 763 CTAGGAATTGTAAATCAGCTTAAGCGA 789

# RESULT 12.

US-09-689-065B-2  
 ; Sequence 2, Application US/09689065B  
 ; Patent No. 6605696

## GENERAL INFORMATION:

; APPLICANT: Pfizer Products, Inc.  
 ; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS AND RELATED METHODS AND  
 ; TITLE OF INVENTION: MATERIALS  
 ; FILE REFERENCE: 3153.00187/PC10589A  
 ; CURRENT APPLICATION NUMBER: US/09/689,065B  
 ; CURRENT FILING DATE: 2000-10-12  
 ; PRIOR APPLICATION NUMBER: US Prov. 60/160,922  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR APPLICATION NUMBER: US Prov. 60/163,858  
 ; PRIOR FILING DATE: 1999-11-05  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 2  
 ; LENGTH: 5445  
 ; TYPE: DNA  
 ; ORGANISM: Lawsonia intracellularis

US-09-689-065B-2

Alignment Scores:  
 Pred. No.: 6,45e-10 Length: 5445  
 Score: 193.00 Matches: 40  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.46% Indels: 0  
 DB: 4 Gaps: 0

US-10-009-823A-1 (1-502) x US-09-689-065B-2 (1-5445)

QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482

Db 3 AACGTAGACATGACGAGAGAAATGGTTATATATGATTATTTCAACGTGTTTCAGATG 62  
Qy 483 AsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlnLeuLysArg 502  
Db 63 AATAGTAATCTGTTACACAGCAGACAAATGCTACAAAGACACTTGAACCTAAAGCGT 122

RESULT 13  
US-09-252-991A-13873/c  
; Sequence 13873, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13873  
; LENGTH: 3129  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13873

Alignment Scores:  
Pred. No.: 1.78e-08 Length: 3129  
Score: 176.00 Matches: 120  
Percent Similarity: 33.21% Conservative: 56  
Best Local Similarity: 22.64% Mismatches: 212  
Query Match: 6.81% Indels: 142  
Gaps: 22

US-10-009-823A-1 (1-502) x US-09-252-991A-13873 (1-3129)  
Qy 20 GlyThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnVal 39  
Db 2609 GGCACGTGTGTCACGCGGTGGCC----- 2586  
Qy 40 ValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGly 58  
Db 2585 -----CAGACCCCTGCG-----GGCAATACCGCGCCCGCAGGGCAGC 2550  
Qy 59 -----ProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal 71  
Db 2549 ACTACCGTGGACGGGTGGCCGACACGCGTGTGTCATCCGACGACGCACTG 2490  
Qy 72 ArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIle 91  
Db 2489 -----CTCAACGCTACCGCGGAGCGCGGCGCAGCACCCTGACCTTGACCGCGC 2442  
Qy 92 GlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGly 111  
Db 2441 AACGGCAACCCGATCGGCGACACCGCCGATGGC-----AGCGGC 2400  
Qy 112 AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131  
Db 2399 AACTGGAGCTTCACGCGCGGTGCGCACTACCCCAACGCGCACCGTGTGTCACGTCACCGCG 2340  
Qy 132 SerArgIleSerAsnAsnProAsnIleLysLysGlnThrLeuGluProIleGlnLeuAsp 151  
Db 2339 AGCGACGCGCGGCAATACCAAGCGCTCCCGCTACCAAG-----ACGGTGGAT 2292  
Qy 152 PheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 171  
Db 2291 TCCTCGCTGCGTGCATCCCGCAGGTGATCCGACGACGCTGCTGCTGATCAGCGGCACC 2232  
Qy 172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu 191

Db 2231 CGCGACGCGCGCAACACCATCATCATCCGATGGCAACGCAACCGATTGGCCAGTGC 2172  
Qy 192 LeuGluSerTyrLysGlyAsnGly-----ThrProIleSerThrSerAsnTyrSer 209  
Db 2171 ACCGCCGACGCGCGGTAACTGCTCTTCACTCCAGCATCCCGCTGCGGATGGCAGC 2112  
Qy 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229  
Db 2111 GTGGTCAACGTGTGGCGCGCAGCCCAAGCAATTCGACATCGCCGCGCGGTGATCACT 2052  
Qy 230 PheAspGly---AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248  
Db 2051 GTGGATGGCGTGGCGCGCGCGCGCGCG-----GTGATC 2019  
Qy 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet--- 267  
Db 2018 GATCCGAGC---AAGCGCAGCAGATAGCGGTACCGCGAGGCGCGCGCGGTGATC 1962  
Qy 268 -----SerGlyThrMet 271  
Db 1961 CTCACCGATGGCGCGGCAACCCGATCGCGCAGGCCCGCGCGCGCGCGCACTGG 1902  
Qy 272 ThrPheSer-----SerAsnGlyGluLeuLysAsnMetThrAlaPheThr 286  
Db 1901 AGTTTACCCCGCGGACCCCGCTGGCCAAACGCGCGGTGATCAACGCGCGTGGCCAGGAC 1842  
Qy 287 ProThrGly-----SerAlaThrLysAspLeuAsnAlaTyrGlnPro 300  
Db 1841 CCGCGCGCGCAATACCGCGGCTCGCGCGCGCTCCGTCGATCGCATCGCC---CCGCG 1785  
Qy 301 AlaProLeuValAsn-----GlyLeuProGlnPheSerAla 312  
Db 1784 GCGCGGTGATCAATCCGAGCAACGAGTGTCTCATCGCGGTGCGCGGAGCGCGGCGC 1725  
Qy 313 AsnPheVal-----GlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327  
Db 1724 ACGGTGATCTCACCGACGCGCAACCGCATCGCGCGCGGTGATCGCGCGCGCGCGCAGC 1665  
Qy 328 IleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThr 347  
Db 1664 GCGAAG-----TGGGCTTTCACGCGCGCGCGCGCGCGCGTGGCCATGGCAGC 1620  
Qy 348 AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla 367  
Db 1619 GTGATCAATGCGCTG-----GCCACGAGCAGCGCGC 1590  
Qy 368 ArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu 387  
Db 1589 GGCACACACAGCAGTCCACACGCGCCACCGCTGCTGCGCGCGCGCGCGCGCGG 1530  
Qy 388 ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal 407  
Db 1529 ATCGATCCG-----AGCAACGCTAGCGTG 1506  
Qy 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427  
Db 1505 ATCGCGGTACCGCGGAGGCTGTCACGCTGATCTCTCACCGCGC-----AAC 1455  
Qy 428 GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447  
Db 1454 GGCACCCCGATCGCGCGCGTCCCGCGGATGCGCGCGCACTGGAGCTTCACGCGCGC 1395  
Qy 448 ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer 467  
Db 1394 ACG-----CCGCTGTCCAAATGGCAGCGTG----- 1371  
Qy 468 ArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerVal 487  
Db 1370 -----GTCAATCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAGC 1320  
Qy 488 ThrAlaAspThrMetLeuGlnLysAla 497  
Db 1319 ACCACGCTGACTCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1290



APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 3047  
LENGTH: 408  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3047

Alignment Scores:  
Pred. No.: 1.04e-09 Length: 408  
Score: 174.00 Matches: 43  
Percent Similarity: 50.00% Conservative: 27  
Best Local Similarity: 30.71% Mismatches: 64  
Query Match: 6.73% Indels: 6  
Gaps: 3  
DB: 4

US-10-009-823A-1 (1-502) x US-09-252-991A-3047 (1-408)  
QY 34 TyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer 53  
Db 3 TTCAGCAATCCCGCGGAGTTCCGCGACGCTACGCCGCCCTCGGTG---CTGGGTTG 59  
QY 54 ThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValaIleGlyThr 73  
Db 60 -----GGCAGCAACCCCGAGCGGCGGCGGTCTCTCGGACGTCCTCGCAG 107  
QY 74 IlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGly 93  
Db 108 ATGTTCAAGCAGGCAACATCATCTCCACCAACAGCGGTGTGACCTGGCCATCAACGGC 167  
QY 94 LysGlyPheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPhe 113  
Db 168 AACGGCTTCTTCGTCAACCAACCAACAGGGCGGATCATCGTCAGCAACACCGCTACCCCGCGGCTACTTC 227  
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133  
Db 228 AATACCGCAACAGCAGGATTTCATCGTCAGCAACACCGCTACCCCTGCGAGGCTATGCC 287  
QY 134 IleSerAsnAsnProAsnIleLysGlnThrLeuGluProIleGlnLeuAsp---Phe 152  
Db 288 GTCGGCGCAACGCCAGTTCAGAACAGCGGTGTCTCCAGACCTCAAGTCGAGCGGCC 347  
QY 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 172  
Db 348 AATCAGGCGCGCAGGCGCCACCTCGAGCATCCAGCGTGTCAACACCTCAACTCGACGCTG 407

RESULT 16  
US-09-206-942-68  
Sequence 68, Application US/09206942  
Patent No. 6432669  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
TITLE OF INVENTION: Molecular Weight Proteins  
FILE REFERENCE: 1038-861 MIS:jb  
CURRENT APPLICATION NUMBER: US/09/206,942  
CURRENT FILING DATE: 1998-12-08  
EARLIER APPLICATION NUMBER: 09/167,568  
EARLIER FILING DATE: 1998-10-07  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 68

LENGTH: 3285  
TYPE: DNA  
ORGANISM: Haemophilus influenzae  
US-09-206-942-68  
Alignment Scores:  
Pred. No.: 3.58e-08 Length: 3285  
Score: 173.50 Matches: 110  
Percent Similarity: 37.30% Conservative: 75  
Best Local Similarity: 22.18% Mismatches: 210  
Query Match: 6.71% Indels: 101  
Gaps: 24  
DB: 4

US-10-009-823A-1 (1-502) x US-09-206-942-68 (1-3285)  
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
Db 1807 AATATTTTCAGGTTTCAATAAAGCAGAGATTACAGTTAAAGAT-----GGTAGTGATT 1860  
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
Db 1861 ACTATTGGTAACCAACCAATAGTCTGCTGCTGCTACTAT----- 1896  
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
Db 1897 ---GCCAAAAAAGTAACCTTT-----AACCAGGTTAAAGAT 1929  
QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108  
Db 1930 TCAAAAATCTCTGCTGACCGGT---CACAGGTGACACACACAGCAAGTAGTGA 1980  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 1981 ACATCGGTAGTAATAACACACACATGACATGACGAGTACCAATAAT---GCCGGCTTAACT 2037  
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlnGluThrLeuGluProIle 148  
Db 2038 ATCGATGCAAAAATGTAAACAGTAAACACATATTACTTCTCAACAGCAGTAGCAGCATC 2097  
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
Db 2098 -----TCTCGCACAGTGGAGAAATACCCTAAACAGGTAACACCAATT 2142  
QY 169 AsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
Db 2143 AACGCCAACCTGTTAAACGTGGAGATAACCGCTCAACACAGGTAGT-----ATC 2190  
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203  
Db 2191 CTAGTGGAAATGAGTCCAGCTCTGCTCTGTAAACACTTACTGCAACCGCGGCTCTT 2250  
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
Db 2251 GCTGTAAAGCAATATTTCG-----GCCAACACC 2277  
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
Db 2278 GTTACTGTTACTGCAAAAT---AGCGGTGATTAACCACTTTGGCGAGCTCTCAATAATAA 2334  
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
Db 2335 GGAACCGAGTGTATACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTCTTCTGT 2394  
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
Db 2395 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCAATCCAAATTCAAAAATTA 2454  
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300  
Db 2455 GCAACACACGCGGAGGCTACGTAAACAGTGTCAACAGGTACATTCGTTGGTACGATTCC 2514  
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320

Db 2515 GGTAAATACGGTAAAT-----GTTACGGCAAAACGCTGGC 2547  
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
Db 2548 GATTAAACAGTTGGGATGGCGCAGAA-----ATTAAATGGCAGCAA 2589  
QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
Db 2590 GGAGCTGCAACCTTAACATCATCATCGGCAATTA-----ACT 2628  
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380  
Db 2629 ACCGAGCTAGTTACACATTAATCTACGCAAGGTCAGGTAAATCTTTTCAGCTCAGGAT 2688  
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398  
Db 2689 GGTAGCGTTGCAGGAAGTATTAAATCGGCCAATGTGACACTAAATACTACAGGCACITTA 2748  
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
Db 2749 ACTACCGTGAAGGGTCAACACATTAATGCAACCGGTACCTTGGTTATTAAACGCCAAA 2808  
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
Db 2809 GACGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGATGGTAAATGCAACCAACGCA 2888  
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
Db 2869 AATGCTCGCGCAGCGGTAAATCGCGACACCTCAAGC-----AGAGTGAACATCACT 2919  
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
Db 2920 GGG-----GATTAAATCAATAATGGAATTA-----AATATC-----ATT 2955  
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
Db 2956 TCMAAAACCGGTATAAACAACCGTACTGTGA-----AAAGCGGTAAATGATGTGAATATAC 3012  
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
Db 3013 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTTGAAGCGAAACGC 3060

## RESULT 17

US-08-038-682-1  
; Sequence 1, Application US/08038682  
; Patent No. 5549857  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682  
; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5116 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-038-682-1

## Alignment Scores:

Pred. No.: 7.39e-08 Length: 5116  
Score: 173.50 Matches: 110  
Percent Similarity: 37.30% Conservative: 75  
Best Local Similarity: 22.18% Mismatches: 210  
Query Match: 6.71% Indels: 101  
DB: 1 Gaps: 24

## US-10-009-823A-1 (1-502) x US-08-038-682-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
Db 3480 AATATTTTCAGGTTTCATTAACGACAGATTACAGCTTAAGAT-----GGTAGTGATTTA 3533  
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
Db 3534 ACTATTGGTAACCAATAGTCTGATGCTACTAAT----- 3569  
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
Db 3570 -----GCCAAAAGTAACCTTT-----AACAGGTTAAGAT 3602  
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108  
Db 3603 TCMAAAATCTCTGCTGACGGT---CACAAAGGTGACACTACACGCAAAAGTGAA----- 3653  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 3654 ACATCGGTGATTAATAACACACTCAAGATAGCAGTGAATAAT---GCCGGCTTAAT 3710  
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
Db 3711 ATCGATGCAAAAATGTAAACAGTAAACACATATTAATCTTCTCACAAGCAGTGAGCATC 3770  
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
Db 3771 -----TCTCGCACAGTGGAGAAATTAACACTAAACAGGTAACCACT 3815  
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
Db 3816 AACGCAACCACTGGTAACGTGAGATACCTGCTCAACAGGTAGT-----ATC 3863  
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProIle 203  
Db 3864 CTAGGTGGAATTTAGTCCAGCTCTGGCTCTGTAAACACTTACTGCAACCGAGGGCGCTCT 3923  
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
Db 3924 GCTGTAAAGCAATATTTTCG-----GGCAACAC 3950  
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
Db 3951 GTTACTGTTACTGCAAAAT---AGCGGTGCAATTAACCACTTTGGCAGGCTCTACAAATAA 4007  
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
Db 4008 GGAACCGAGAGTGAACCACTTCAAGTCAATFCAGCGCATATTCGGTTCCTTCTGGT 4067  
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
Db 4068 GGCACAGTAGAGTTAAGCAACCGAAAGTTTAACTCAATCCATTCATAAATAA 4127



QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
Db 3951 GTTACTGTTACTGCAAAAT---AGCGGTGCATTAACACCTTTGGCAGGCTCTACAAATAA 4007  
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
Db 4008 GGNACCGAGGTATACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTCTGGT 4067  
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGlyLeuLys 280  
Db 4068 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCAATCCAAATTCAAAATAA 4127  
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300  
Db 4128 GMAACACAGCGGAGGCTAACGTAAACAGTCAACAGGTACAAATGGTGGTACGATTTC 4187  
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
Db 4188 GGTAAATACGGTAAAT-----GTTACGCAACCGCTGGC 4220  
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
Db 4221 GATTTAACAGTTGGGAATGGCCAGAA-----ATTAATGCGACAGAA 4262  
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
Db 4263 GGAGCTGCAACCTTAACATCATCGGCGCAATTA-----ACT 4301  
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgATyrsSerGlnAsp 380  
Db 4302 ACCGAAGCTAGTTTACATTAACCAAGGTCAGGTAAATCTTTCAGCTCAGAT 4361  
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398  
Db 4362 GGTAGCGTTGCAAGGAAGTATTAAATCCGCCAATGTGACACTAAATACTACAGGCACCTTA 4421  
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
Db 4422 ACTACCGTGAAGGTTCAACATTAATGCAACCGGTACCTTGGTTATTATACGCAAA 4481  
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
Db 4482 GACGCTGAGCTAAATGCGCGAGCATTTGGTAAACACACAGATGGTAAATGCAACCAACGCA 4541  
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
Db 4542 AATGGCTCCGCGAGCGTAATCGGACACCTCAAGC-----AGAGTGAACATCACT 4592  
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
Db 4593 GGG-----GATTTAATCACAATAAATGGATTA-----AATATC---ATT 4628  
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
Db 4629 TCAAAAACGGTATAACACACCGTACTGTTA---AAAGCGTTAAATGATGTAATATC 4685  
QY 487 ValThrAlaAspThrMetLeuGlnLysAlaLeuLysArg 502  
Db 4686 ATTCAACCGGTATAGCAAGCGTAGATGAATTAATGAAAGCAACGC 4733

## RESULT 19

US-08-530-198-1  
; Sequence 1, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/530,198

FILING DATE: 13-DEC-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERKSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: JWB-1186

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5116 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-10-009-823A-1

Alignment Scores:

Prod. No.: 7.39e-08 Length: 5116

Score: 173.50 Matches: 110

Percent Similarity: 37.30% Conservative: 75

Best Local Similarity: 22.18% Mismatches: 210

Query Match: 6.71% Indels: 101

DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-530-198-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49  
Db 3480 AATATTTTCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATTTA 3533  
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
Db 3534 ACTATTGGTACCAACCAATAGTCTGCTGCTACTAAT----- 3569  
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
Db 3570 ---GCCAAAAAAGTAACCTTT-----AACCAAGTTAAAGAT 3602  
QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrIleuGluAspLysValHisThr 108  
Db 3603 TCAAAAATCTCTGTCGACGGT---CACAAAGGTGACACTACACAGCAAGTGA--- 3653  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 3654 ACATCCGGTAGTAAATAACACACTGAAGTAGCAGTGAATAAT---GCCGGCTTAAT 3710  
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
Db 3711 ATCGATGCAAAAATGTAAACAGTAGTAACCAATAATTACTTCTCAAAAGCAGTGAAGCATC 3770  
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
Db 3771 -----TCTGCGACAAAGTGGAGAAATACCCTAATAACAGGTACACCAT 3815  
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
Db 3816 AACGCAACCACTGGTAACTGAGATACCGCTCAAAACAGGTAGT-----ATC 3863

|      |    |                                                               |      |
|------|----|---------------------------------------------------------------|------|
| 189  | QY | PheAlaLeuLeuGluSerTrpTyrGlyAsnGlyThr-----ProProfile           | 203  |
| 3864 | Db | CTAGGTGAAATGTAGTCCAGCTCTGCTCTGTAAACACTTACTGCAACGAGGGCGCTCTT   | 3923 |
| 204  | QY | SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnIysAsnSer  | 223  |
| 3924 | Db | GCTGTGAAGCAATATTCG-----GGCAACACC                              | 3950 |
| 224  | QY | HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu  | 243  |
| 3951 | Db | GTTACTGTTACTCAAAAT---ACGGTGGATTAACCACTTTGGCAGGCTCTACAAATTAAA  | 4007 |
| 244  | QY | TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla  | 263  |
| 4008 | Db | GGAACCGAGAGTGTACCACTTCAAGTCAATCAGCGCATATCGCGGTACGATTTCTGGT    | 4067 |
| 264  | QY | GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys      | 280  |
| 4068 | Db | GCCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCAATCCAAATTCAAAAATTAAA | 4127 |
| 281  | QY | AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro  | 300  |
| 4128 | Db | GCAACACAGCGAGGCTAACGTAACAAAGTGCACACAGGTACAAATGTGGTGCAGATTTC   | 4187 |
| 301  | QY | AlaProIleuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln | 320  |
| 4188 | Db | GGTAATACGGTAAAT-----GTTACGGCAACACGGTGGC                       | 4220 |
| 321  | QY | ProLeuThrLeuAspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAla     | 340  |
| 4221 | Db | GATTACAGTTGGGATGCGCGAA-----ATTATGCGACAGAA                     | 4262 |
| 341  | QY | SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr     | 360  |
| 4263 | Db | GGAGCTTCAACCTTAACTACATCATCGGCCAAATTA-----ACT                  | 4301 |
| 361  | QY | SerSerGlyAsnSerThrAlaAspAsnGlySerSerSerThrArgTyrSerGlnAsp     | 380  |
| 4302 | Db | ACCGNAGTAGTTCACACATTCTTCAGCAAGGTCAGTAAATCTTCAGCTCAGAT         | 4361 |
| 381  | QY | GlyTyrProGlnIleAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu   | 398  |
| 4382 | Db | GSTAGCGTTCAGAGAAATTAATTCGCGCAATGTGCACATAAATACTACAGGCACCTTA    | 4421 |
| 399  | QY | GlnGlyLys-----TyrSerAsnSerGln                                 | 406  |
| 4422 | Db | ACTACCGTGAAGGTTCAACATTATGCAACCGGTCCTTGGTTATTATACGCAAAA        | 4481 |
| 407  | QY | ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg     | 426  |
| 4482 | Db | GACGCTCAGCTAAATGCGCAGCATTTGGGTACACACAGTGGTAAATGCAACCAACGCA    | 4541 |
| 427  | QY | GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro  | 446  |
| 4542 | Db | AATGGCTCCGGCAGCGTAAATCGCGACCACTCAACG-----AGAGTGAACATCACT      | 4592 |
| 447  | QY | GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet  | 466  |
| 4593 | Db | GGG-----GATTATTAATCAATAAATGGATTA-----AATATC---ATT             | 4628 |
| 467  | QY | SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer     | 486  |
| 4629 | Db | TCRAAAAACGGTATTAACACCGTACTGTGTA---AARGCGTTAAATTAATGATGAATAC   | 4685 |
| 487  | QY | ValThrThrAlaAspThrMetLeuGlnIysAlaLeuGluLeuLysArg              | 502  |
| 4686 | Db | ATTCAACCGGTATAGCAACGGTAGATGAAGTAATTGAAGCAACCG                 | 4733 |

## RESULT 20

RESOL 20  
US-08-469-880-1

; Sequence 1, Applica

; Patent No. 5876733

**; GENERAL INFORMATION:**

1000

```

APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
FILE NO.: 08-067933-1
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MJS:vq
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5116 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-1

Alignment Scores:
Pred. No.: 7,39e-08 Length: 5116
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: Gaps: 24

US-10-009-823A-1 (1-502) x US-08-469-880-1 (1-5116)

Qy 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSeSere
Db 3480 AATATTTCAGTTTCAATAAAGCAGAGATACAGCTAAAGT-----GGGT
Qy 50 AlaAlcGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGly
Db 3534 ACATTTGGTAACACCACCAATAGTCTGTGTTGGTACTAAAT
Qy 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSeSere
Db 3570 ---GCCAAAAGAAGTAAACCTTT-----AACCAAG
Qy 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGlnAspLysVal
Db 3603 TCAAAAATCTCTCTGACCGGT---CACAAAGTGACACATACACAGCAAAGT
Qy 109 ArgAlaGlyAsnPheArqPheThrGlnAspGlyPheLeuAsnAspProSeSere

```

## RESULT 20

RESUL 20  
US-08-469-

; Sequence

; Patent N  
 : Governat

**GENERAL**

1

\_\_\_\_\_



3480 ATATTTCAGGTTTCATTAAGACAGAGATTACAGCTAAGAT-----CGTAGTGATTTA 3533  
50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
3534 ACTATTGGTAACCAATAAGTCTGCTGATGCTACTAAT----- 3569  
69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
3570 ---GCCAAAAAGTAACCTTT-----AACAGGTTAAAGAT 3602  
89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyThr 108  
3603 TCAAAATCTCTGCTGACGCT---CACAGGTCACACTACACAGCAAGTGGAA----- 3653  
109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
3654 ACATCCGGTAGTAAATACACACTGAAGTAGAGTAGCAGTGACAAATAAT---GCCGCTTAATC 3710  
129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
3711 ATCATGCCAAAAATGTACAGTAAACAAATAATTACTTCTCACAAAGCAGTGAGCATC 3770  
149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
3771 -----TCTGGCACAAGTGGAGAAATTACCCTAAACACAGGTACACACCAT 3815  
169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyThr 188  
3816 AACCAACCACTGTAACGTGAGATACCGCTCAACACAGTAGT-----ATC 3863  
189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203  
3864 CTAGGTGGAATTGAGTCCAGCTCTGCTGCTGTAACTTACTGCAACCGGCGCTCTT 3923  
204 SerThrSerAsnTyThrValAlaGlnProMetArgValTyThrAspGlnGlyAsnSer 223  
3924 GCTGTAAAGCAATATTCG-----GGCAACACC 3950  
224 HisAspIleThrValTyPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
3951 GTTACTGTACTGCAAT---AGCGGTGATTAACCACTTGCGAGGCTCTACAAATATAA 4007  
244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaLysThrAspSerAla 263  
4008 GGAACCGAGAGTAAACCACTCAAGTCAATCAGCGCATATCGCGGTACGATTCTCGT 4067  
264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
4068 GGCACAGTAGAGTTAAGCAACCGAAGTTTAACTCACTCAATCCAAATTCAAAAATAA 4127  
281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300  
4128 GCAACACACAGCGAGGCTAACGTAACTGCAACAGGTACAAATGCTGTGATTCACATTCC 4187  
301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
4188 GGTAAATACCGTAAAT-----GTTACGGCAACCGCTGGC 4220  
321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
4221 GATTAAACAGTTGGGAATGGCGAGAA-----ATTAATGCGACAGAA 4262  
341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
4263 GGAGCTGCACACCTTAACCTACATCATCGGCGAAATTA-----ACT 4301  
361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyThrSerGlnAsp 380  
4302 ACCGAAGCTAGTTTACACATTAATCTACGCCAAGGCTCAGGTAAATCTTTCAGCTCAGAT 4361  
381 GlyTyThrProGlnGlyAspLeu-----ValAspValThrIleThrSerGlyLysLeu 398  
4362 GGTAGGTTGCGAGGAAGTATTAAATGCGCCCAATGTGACACTAAATACTACAGGCACCTTA 4421

399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
4422 ACTACCGTGAAGGTTCAACATTAATGCAACGCGGTACCTTGGTTATTAAACGCAAAA 4481  
407 ValValAspPheTyAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
4482 CACGCTGAGCTAAATGCGCAGCATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4541  
427 GlnGlyAsnAsnHisTyThrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
4542 AATGCTCCGCGACGGTAATCGCACACCTCAAGC-----AGAGTGACATCACT 4592  
447 GlyThrSerAsnTyGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
4593 GGG-----GATTTAATCACATAAATGGATTA-----AATATC---ATT 4628  
467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
4629 TCMAAAACCGTATAAACACCGTACTGTTA---AAAGCGTTAAATGATGTGAATAC 4685  
487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
4686 ATTCAACCGGTTATAGCAACGCGTAGATGAAGTAATGAAGCAACGCG 4733

RESULT 22  
US-08-617-697-1  
; Sequence 1, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 11  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-557  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-617-697-1

## Alignment Scores:

Pred. No.: 7,398-08 Length: 5116  
 Score: 173.50 Matches: 110  
 Percent Similarity: 37.30% Conservative: 75  
 Best Local Similarity: 22.18% Mismatches: 210  
 Query Match: 6.71% Indels: 101  
 DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-617-697-1 (1-5116)

QY 30 AsnThrIleGlyTyrGlnGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3480 AATATTTCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATTTA 3533  
 Db ||||| : : : : : ||||| : : : : :  
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3534 ACTATTGGTAACCAATAGTCTGATGCTAATA----- 3569  
 Db ||||| : : : : : ||||| : : : : :  
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3570 ---GCCAAAAAATTAACCTTT-----AACCGATTAAAGAT 3602  
 Db ||||| : : : : : ||||| : : : : :  
 QY 89 LeuAlaIleGlyGlyGlyPheGlnValThrLeuGluAspLysValHisThr 108  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3603 TCAAAAATCTCTGCTGACGGT---CACAAAGTGACACTACACAGCAAGTGAA----- 3653  
 Db ||||| : : : : : ||||| : : : : :  
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3654 ACATCCGGTAGTAATAACACACTGAAGATGACGTGACATATAT---GCCGGCTTAAC 3710  
 Db ||||| : : : : : ||||| : : : : :  
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlnThrLeuGluProIle 148  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3711 ATCGATGCAAAAATGTAAACAGTAACCAATATTAATCTCTCAAAAGCAGTGAGCATC 3770  
 Db ||||| : : : : : ||||| : : : : :  
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3771 -----TCTGCACAGTGGAGAAATTAACCAATTAACCAAGTACCAACCAT 3815  
 Db ||||| : : : : : ||||| : : : : :  
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3816 AACGCAACCACTGGTAACCTGAGATAACCGCTCAACACAGTAGT-----ATC 3863  
 Db ||||| : : : : : ||||| : : : : :  
 QY 189 PheAlaLeuLeuGluSerThrLysGlyAsnGlyThr-----ProProIle 203  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3864 CTAGGTGAATGTAGTCCAGCTCTGGCTCTGTAACTTACTGCAACCGAGGCGCTCTT 3923  
 Db ||||| : : : : : ||||| : : : : :  
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3924 GCTGTAGCAATATTTCG-----GGCAACACC 3950  
 Db ||||| : : : : : ||||| : : : : :  
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
 Db ||||| : : : : : ||||| : : : : :  
 QY 3951 GTTACTGTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAAATAAA 4007  
 Db ||||| : : : : : ||||| : : : : :  
 QY 244 TyrLeuValAlaAsnProSerGlyAspGlySerAlaAlaSerGlyThrAspSerAla 263  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4008 GGAACCGAGAGTGTAAACCACTTCAAGTCATCAGCGGATTCGGGTAGCATTTCTGGT 4067  
 Db ||||| : : : : : ||||| : : : : :  
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4068 GGCACAGTAGAGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAATTCMAAAATATA 4127  
 Db ||||| : : : : : ||||| : : : : :  
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4128 GCAACACAGCGGCGGTACGTAAACAGTGCAACAGGTACCAATTGGTGTAGCATTTCC 4187  
 Db ||||| : : : : : ||||| : : : : :  
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4188 GGTAATACGGTAAAT-----GTTACGCAACACGCTGCGC 4220  
 Db ||||| : : : : : ||||| : : : : :  
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4221 GATTTAACAGTGGGAATGCGGAGAA-----ATTATGCGCAGAA 4262  
 Db ||||| : : : : : ||||| : : : : :

QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4263 GGAGCTGCAACCTTAACCTATCATCGGCAAAATTA-----ACT 4301  
 Db ||||| : : : : : ||||| : : : : :  
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4302 ACCGAAGCTAGTTTCACACATTACTTCGCCCAAGGTCAGGTAATCTTTTCAGCTCAGGAT 4361  
 Db ||||| : : : : : ||||| : : : : :  
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4362 GGTAGCGTTTCAGGAAGTATTAAATGCCGCCAATGTGACACTAAATACTACAGCACTTTA 4421  
 Db ||||| : : : : : ||||| : : : : :  
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4422 ACTACCGTGAAGGGTTCAAAACATTAAATGCAACCGGTACCTTGGTTATTAAACGCAAAA 4481  
 Db ||||| : : : : : ||||| : : : : :  
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4482 GAGCTGAGCTAAATGGCGCAGCATTTGGTAACCAACAGTGGTAATTCACCAACCAACGA 4541  
 Db ||||| : : : : : ||||| : : : : :  
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4542 AATGGCTCCGCGAGCGTAAATCGCAACCTCAAGC-----AGAGTGAACATCACT 4592  
 Db ||||| : : : : : ||||| : : : : :  
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4593 GGG-----GATTTAATCAATAATGGAATTA-----AATATC---ATT 4628  
 Db ||||| : : : : : ||||| : : : : :  
 QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4629 TCAAAAACCGGTATAAACACCCGTACTGTTA---AAAGCGGTAAATTTGATGTGAATAC 4685  
 Db ||||| : : : : : ||||| : : : : :  
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
 Db ||||| : : : : : ||||| : : : : :  
 QY 4686 ATTCAACCGGTATAGCAAGCGGTAGATGAAGTAATTTGAACGCAACCGC 4733  
 Db ||||| : : : : : ||||| : : : : :

## RESULT 23

US-08-719-641-1  
 ; Sequence 1, Application US/08719641  
 ; Patent No. 6218141  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; STREET: Bldg. 1  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 16-SEP-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9205704.1  
 ; FILING DATE: 16-MAR-1992  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 1038-625  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5116 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-719-641-1

# Alignment Scores:

Pred. No.: 7,398-08 Length: 5116  
 Score: 173.50 Matches: 110  
 Percent Similarity: 37.30% Conservative: 75  
 Best Local Similarity: 22.18% Mismatches: 210  
 Query Match: 6.71% Indels: 101  
 DB: 3 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-719-641-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
 DB 3480 AATATTTCAGGTTTCAATAAAGACGAGATTACAGCTAAAGAT-----GGTAGTGATTTA 3533  
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
 DB 3534 ACTATTGGTAACACCAAGTAGTGCTGATGTTACTAAT----- 3569  
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
 DB 3570 ---GCCAAAAGTAACCTTT-----AACCCAGGTAAAGAT 3602  
 QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108  
 DB 3603 TCAAAATCTCTGCTGACGGT---CACAGGTGACACTACACGAAAGTGGA----- 3653  
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
 DB 3654 ACATCCGGTAGTAATAACACACTGAAGATAGCAGTACCAATAT---GCCGGCTTAAC 3710  
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
 DB 3711 ATCGATGCAGAAAATGTACAGTAACACACATATTACTTCTCACAAAGCAGTGAGCATC 3770  
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
 DB 3771 -----TCTGCAGCAAGTGGAGAAATTACCACTAAACACAGGTACACCAATT 3815  
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
 DB 3816 AACGCAACACTGGTAACGTGAGATAACCCCTCAAACAGTAGT-----ATC 3863  
 QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203  
 DB 3864 CTAGGTGAAATAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGGGCGCTCT 3923  
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsnSer 223  
 DB 3924 GCTGTAAACAAATTTTCG-----GCCACACC 3950  
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerThrPheGlu 243  
 DB 3951 GTTACTGTACTGCAAAAT---AGCGGTGCAATTAACCACTTTGGCAGGCTCTACAAATTA 4007  
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
 DB 4008 GGAACCGAGAGTGTACCACTTCAAGTCAATCAGCGGATATCGCGGTAGCATTCTTG 4067

QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
 DB 4068 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAAATTCAAAATTA 4127  
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300  
 DB 4128 GCAACAACACGCGGAGGCTAACGTAAAGTCAACAGCTCAATTTGGTGTACGATTTC 4187  
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
 DB 4188 GGTAAATACGGTAAAT-----GTTACGGCAACCGTGGC 4220  
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
 DB 4221 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTAATGCGACAGAA 4262  
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
 DB 4263 GGAGCTGCACCTTAACCTACATCATCGGCAAAATTA-----ACT 4301  
 QY 361 SerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380  
 DB 4302 ACCGAAGCTAGTTCCACACATTACTTACGCAAGGTCAGGTAAATCTTTTCAGCTCAG 4361  
 QY 381 GlyTyrProGlnGlyAspLeu-----ValaspValThrIleThrSerGluGlyLysLeu 398  
 DB 4362 GGTAGCGTTGCAAGAGTATTAAATGCCCAATGTGACACTAAATACTACAGGCCTTTA 4421  
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
 DB 4422 ACTACCGTGAAGGTTCAAAACATTAAATGCAACGCGGTACCTTGGTTATTAAACGCAAA 4481  
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
 DB 4482 GACGCTGAGCTAAATGCGCAGCAGCATTCGGTAAACACACAGCTGTAATAATCAACCA 4541  
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
 DB 4542 AATGGCTCGGACGCGTAATCGCAGCAACCTCAAGC-----AGAGTGAACATCACT 4592  
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
 DB 4593 GGG-----GATTTAATCAATAAATGCGATTA-----AATATC---ATT 4628  
 QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
 DB 4629 TCAAAAACGCTATAAACACCGTACTGTTA---AAAGGCGTTAAATTTGATGTGAATAC 4685  
 QY 487 ValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
 DB 4686 ATTCACCGCGGTATAGCAAGCGTAGATGAAGTAATTTGAAGCGGAACGC 4733

## RESULT 24

US-09-206-942-66  
 Sequence 66, Application US/09206942  
 Patent No. 6432869  
 GENERAL INFORMATION:  
 APPLICANT: Loosmore, Sheena M.  
 APPLICANT: Yang, Yan-ping  
 APPLICANT: Klein, Michel H.  
 TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 TITLE OF INVENTION: Molecular Weight Proteins  
 FILE REFERENCE: 1038-861 M1S:jfb  
 CURRENT APPLICATION NUMBER: US/09/206,942  
 CURRENT FILING DATE: 1998-12-08  
 EARLIER APPLICATION NUMBER: 09/167,569  
 EARLIER FILING DATE: 1998-10-07  
 NUMBER OF SEQ ID NOS: 95  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 66  
 LENGTH: 5116  
 TYPE: DNA  
 ORGANISM: Haemophilus influenzae



INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9171 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-038-682-5

## Alignment Scores:

Pred. No.: 1.92e-07  
 Score: 173.50  
 Percent Similarity: 37.30%  
 Best Local Similarity: 22.18%  
 Query Match: 6.71%  
 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-038-682-5 (1-9171)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
 DB 3430 AATATTTTCAGGTTTCATTAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATTTA 3483  
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
 DB 3484 ACTATTGGTAAACCAATAGTCTGTGATGCTACTAAT----- 3519  
 QY 69 GlySerValAlaArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
 DB 3520 ---GCCAAAAGTAACCTTT-----AACAGGTTTAAAGAT 3552  
 QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108  
 DB 3553 TCATAAATCTCTGCTGAGCGT---CACAAAGTGACACTACACAGCAAAAGTGGAA----- 3603  
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
 DB 3604 ACATCCGGTAGTATTAACACACTGAGTAGAGTACAGTACATTAAT---GCCGGCTTAAC 3660  
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
 DB 3661 ATCGATGCAAAATATGTAACAGTAAACACAAATATTACTTCTCACAAAGCAGTGAGCATC 3720  
 QY 149 GlnLeuAspPheAsnAsnProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
 DB 3721 -----TCTGCACACAGTGGAGAAATTACCACTAAACAGGTACACCAT 3765  
 QY 169 AsnAlaValAlaAsnLeuGlyAsnSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
 DB 3766 AACGCACACCACTGTGTAACGTGAGATAACCGCTCAAAACAGGTAGT-----ATC 3813  
 QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProProIle 203  
 DB 3814 CTAGGTGGAATTGATCCAGCTCTGGCTCTGTACTTACTCAACCGAGCGGCTCTT 3873  
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
 DB 3874 GCTGTAGCAATATTTTCG-----GGCAACAC 3900  
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
 DB 3901 GTTACTGTATTCGAAT---AGCGTGTCATTACCACTTTGGCAGGCTCTACATTA 3957  
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
 DB 3958 GGAACCGAGAGTGAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTG 4017  
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
 DB 4018 GCACAGTAGAGGTTAAGAACCAAGATTATACCACTCAATCCCAATTCAAAATTA 4077  
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrPro 300

DB 4078 GCAACAACAGCGAGGCTAACGTAACAGTGCAACAGGTACAAATTCGTTGATCGATTCC 4137  
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
 DB 4138 GGTAATACGGTAAAT-----GTTACGGCAACAGCTGGC 4170  
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
 DB 4171 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTATGGCAGCAAA 4212  
 QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
 DB 4213 GGAGCTGCAACCTTAACATCATCGGCAAAATTA-----ACT 4251  
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380  
 DB 4252 ACCGAAGCTAGTTCACACATTCTTCAGCAAGGTCAGGTAATATCTTCAGCTCAGGAT 4311  
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398  
 DB 4312 GGTAGCGTTGCAGGAAGTATTAAATGCGCAATGTGACACTAAATACTACAGCACTTTA 4371  
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
 DB 4372 ACTACCGTGAAGGTTCAAAACATTAATGCAACGAGGTACCTTGTTTATTACGCAAA 4431  
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
 DB 4432 GACCTGAGCTAAATGCGCAGCATTTGGTAAACACACAGTGTGTAATCAACCAACGCA 4491  
 QY 427 GluclyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
 DB 4492 AATGGCTCCGCGACGGTAATCGGCAACCTCAAGC-----AGAGTGAACATCACT 4542  
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
 DB 4543 GGG-----GATTTAATCAACAATAATGGATTA-----AATATC---ATT 4578  
 QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
 DB 4579 TCAAAAACCGGTATAAACACCGCTACTGTTA---AAAGCGCTTAAATTTGATGGAATAC 4635  
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
 DB 4636 ATTCACCGGCTATAGCAACGCTAGATGAAGTAATTTGAACGCAACGC 4683

## RESULT 26

US-08-302-832-5  
 Sequence 5, Application US/08302832  
 Patent No. 5603938

## GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J  
 TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd.  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/302,832  
 FILING DATE: 16-SEP-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1  
 FILING DATE: 16-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US pct/us93/02166  
 FILING DATE: 16-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Berkstresser, Jerry W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 1038-404  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9171 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-302-832-5

## Alignment Scores:

Pred. No.: 1,92e-07 Length: 9171  
 Score: 173.50 Matches: 110  
 Percent Similarity: 37.30% Conservative: 75  
 Best Local Similarity: 22.18% Mismatches: 210  
 Query Match: 6.71% Indels: 101  
 DB: 1 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-302-832-5 (1-9171)

QY 30 AsnThrLeuGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
 DB 3430 AATATTTCAGGTTTCAATAAGCAGACGATTACAGCTAAAGAT-----GGTAGTGATTTA 3483  
 QY 50 AlaIleGlySerThrGlySer-----GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
 DB 3484 ACTATTGGTAACCAATAGTGTGATGTTACTAAT----- 3519  
 QY 69 GlySerValArgThrIlePheThrGlnGlnValAlaPheGluProGlyAsnSerValThrAsp 88  
 DB 3520 --GCCAAAAGTAACCTTT-----AACCAAGTTAAAGAT 3552  
 QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108  
 DB 3553 TCAAAAATCTCTGCTGACGGT---CACAGGTGACACTACACAGCAAGTGGA----- 3603  
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
 DB 3604 ACATCCGGTAGTAATAACAACACTGAAGTAGCAGTGACAAATAT---CCCGCTTAATC 3660  
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
 DB 3661 ATCGATGCAAAAATGTAAACAGTAACCAACATATATTCTTCACAAAGCAGTGAGCATC 3720  
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
 DB 3721 -----TCTGGCAGTGGAGAAATTAACCACTAAACAGGTCAACACCAT 3765  
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
 DB 3766 AAGCAACCACTGGTAACGTGGAGATAACCGCTCAACACAGGTAGT-----ATC 3813  
 QY 189 PheAlaLeuLeuGluSerThrLysGlyAsnGlyThr-----ProIle 203  
 DB 3814 CTAGGTGGAAATGAGTCCAGCTCTGCTCTGTAAACACTTACTGCAACCGGCGCTCTT 3873  
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
 DB 3874 GCTGTAAAGCAATATTTCG-----GGCAACACC 3900  
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
 DB 243 ----- 243

DB 3901 GTTACTGTTTACTGCAAAAT---AGCGGTGCATTAACCACTTTGCGAGGCTCTACAAATAAA 3957  
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
 DB 3958 GGAACCGAGAGTGAACCACTTCAAGTCAATCAGCGGATATCGGGGTACGATTTCTGGT 4017  
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
 DB 4018 GGCACAGTAGAGGTTAAAGCAACCGAAGCTTTAACCACTCAATCCATTTCAAAATATAA 4077  
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrPheGlnPro 300  
 DB 4078 GCACAAACAGCGGAGCTAACCTAACAGGTGCAACAGGTACAAATTTGGTGGTACGATTTCC 4137  
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
 DB 4138 GGTAAATACGGTAAAT-----GTTACGCAACCGCTGCG 4170  
 QY 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
 DB 4171 GATTTAAACAGTTGGCAATGCGGCAGAA-----ATTAATGCGACAGAA 4212  
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360  
 DB 4213 GCGAGCTGCACCTTAACCTACATCATCGGGCAATTA-----ACT 4251  
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgTyrSerGlnAsp 380  
 DB 4252 ACCGAAGCTAGTTTACACATTTACTTACGCCAAGGTCAGGTAAATCTTTTCAGCTCAGAT 4311  
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398  
 DB 4312 GGTAGCGTTGAGGAAGTATTAAATGCCGCCAATGTGACACTAAATACACAGGCACCTTTA 4371  
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406  
 DB 4372 ACTACCGTGAAGGTTTCAACATTAATGCAACCGGTACCTTGGTTATTAAACGCAAA 4431  
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426  
 DB 4432 GACGCTGAGTAAATGCGCGACGATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4491  
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
 DB 4492 AATGGCTCCGCGACGCGTAATCGCGACACACCTCAAGC-----AGAGTGAACATCACT 4542  
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
 DB 4543 GGG-----GATTTAAATCACAATAAATGGATTA-----AATATC---ATT 4578  
 QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
 DB 4579 TCAAAAACCGGTATAACACCGTACTGTTA---AAAGCGGTAAATGTATGTGAATATAC 4635  
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
 DB 4636 ATTCAACCGGTTAGCAAGCGTAGATGAAGTAATTGAACGCAACGCGC 4683

## RESULT 27

US-08-530-198-5  
 ; Sequence 5, Application US/08530198  
 ; Patent No. 5869065  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BARENKAMP, STEPHEN J  
 ; APPLICANT: ST. GENE III, JOSEPH W  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
 ; NUMBER OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; STREET: Bldg. 1  
 ; CITY: Arlington



NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd.  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg. 1  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469,880  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9205704.1  
 FILING DATE: 16-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US93/02166  
 FILING DATE: 16-MAR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/302,832  
 FILING DATE: 16-SEP-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Berkstresser, Jerry W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9171 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-469-880-5

Alignment Scores:  
 Pred. No.: 1.92e-07 Length: 9171  
 Score: 173.50 Matches: 110  
 Percent Similarity: 37.30% Conservative: 75  
 Best Local Similarity: 22.18% Mismatches: 210  
 Query Match: 6.71% Indels: 101  
 DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-469-880-5 (1-9171)

```

Qy 129 LeuMetGlySerArgIleSerAsnProAsnIleGlySerGluThrLeuGluProIle 148
Db 3661 ATCGATGCAAAAAATGTAACAGTAACAAATATTTACTTCTCACAACGAGTGACATC 3720
Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 3721 -----TCTCGGACAGTGGAGAAATACCACTAAACACAGGTCAACCAT 3765
Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 3766 AACGCAACCACTGGTAACCGTGAGATAACCGCTCAACACAGGTAGT-----ATC 3813
Qy 189 PheAlaLeuLeuGluSerThrLysGlyAsnGlyThr-----ProIle 203
Db 3814 CTAGGTGGAAATGAGTCCAGCTCTGCTCTGTAAACACTTACTGCAACCGAGCGCTCTT 3873
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 3874 GCTGTAAAGCAATATTTTCG-----GGCAACACC 3900
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 3901 GTTACTGTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAAATAA 3957
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 3958 GGAACCGAGAGTGTAAACCACTTCAAGTCAATCAGCGGATATCGCGGTACCAATTTCTGT 4017
Qy 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyLeuLeuLys 280
Db 4018 GGCACAGTAGAGTTAAAGCAACCGAAAGTTTAACTCACTCAATCCAAATCAAAATAA 4077
Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
Db 4078 GCACAACAGCGGAGGTAAAGTAAAGTCAACAGGTCAACATTTGGTGGTACGATTTCC 4137
Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
Db 4138 GGTAAATACGTAAT---GTTACGGCAACCGTGGC 4170
Qy 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
Db 4171 GATTTACAGTTGGATGGCGCAGAA-----ATTAATGCCACAGAA 4212
Qy 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 4213 CGAGCTGCAACCTTAACCTACATCATCGGCAAAATTA-----ACT 4251
Qy 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
Db 4252 ACCGAAGCTAGTTTCAACATTTACTTCAAGCAAGGTCAGGTAAATCTTTCAGCTCAGAT 4311
Qy 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 4312 GGTAGCGTGGTGGAGAGTATTAATGCCCAATGTGACACTAAATACATACAGGCACTTTA 4371
Qy 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4372 ACTACCTGAGGTTCAAAACATTAATGCAACCGGTACCTTGGTTATTAAACGCAAA 4431
Qy 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
Db 4432 GACCTGAGCTAAATGGCGGAGCATTTGGTAAACACAGTGGTAAATGCAACCAACGCA 4491
Qy 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4492 AATGGCTCCGCGAGCGTAATCGCGCAACCACTCAAGC-----AGAGTGAACATCACT 4542
Qy 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 4543 GGG-----GATTTAATCACAATAAATGGATTA-----AATATC---ATT 4578
Qy 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486

```



Db 4372 ACTACGCTGAAGGGTTCAAAACATTAAATGCAACACCGGCTACCTGGTTATTAAACGCAAAA 4431  
 QY 407 ValValAspPheTyrAsnIleProLeuAlaAArgPheThrSerGluAspGlyLeuArgArg 426  
 Db 4432 GACGCTGACCTAAATGGCCACCATTTGGGTACCAACACAGCTGGTAAATGCAACACCGCA 4491  
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446  
 Db 4492 AATGGCTCCGGCAGCGTAAATCCGCAACACCTCAAGC-----AGAGTGAACATCACT 4542  
 QY 447 GlyThrSerAsnTyrGlyLysSerValAsnGlnLeuGluThrSerAsnValAspMet 466  
 Db 4543 GGG-----GATTAAATCAATAATGGATT-----AATATC---ATT 4578  
 QY 467 SerAsgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
 Db 4579 TCAAAAACCGGTATAAACAACCGTACTGTGA---AAAGCGGTAAATAATGTGATGTAATAC 4635  
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502  
 Db 4636 ATTCAACCGGTATAGCAAGCGGTAGATGAAGTAATTGAAGCGCAACCGC 4683

## RESULT 30

US-08-617-697-5  
 ; Sequence 5, Application US/08617697  
 ; Patent No. 5977336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Shoemaker and Mattare, Ltd.  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; STREET: Bldg. 1  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/617,697  
 ; FILING DATE: 01-APR-1996  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 05-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Berkstresser, Jerry W  
 ; REGISTRATION NUMBER: 22,651  
 ; REFERENCE/DOCKET NUMBER: 1038-557  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 415-0810  
 ; TELEFAX: (703) 415-0813  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9171 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-617-697-5  
 Alignment Scores: 1.92e-07 Length: 9171  
 Pred. No.: 173.50 Matches: 110  
 Score:

Percent Similarity: 37.30% Conservative: 75  
 Best Local Similarity: 22.18% Mismatches: 210  
 Query Match: 6.71% Indels: 101  
 DB: 2 Gaps: 24  
 US-10-009-823a-1 (1-502) x US-08-617-697-5 (1-9171)  
 QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49  
 Db 3430 AATATTTCAGGTTTCATAAAGCAGAGATTACAGTTAAAGAT-----GGTAGTGATTTA 3483  
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68  
 Db 3484 ACTATTGGTAACCAATAATAGTGTGATGTTACTAAT----- 3519  
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88  
 Db 3520 ---GCCAAAAAGTAACCTTT-----AACAGGTTAAAGAT 3552  
 QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108  
 Db 3553 TCAAAAATCTCTGCTGACCGT---CACAAAGGTGACACTACACAGCAAAAGTGAA----- 3603  
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
 Db 3604 ACATCCGGTAGTAATAAACAACACTGAAGTAGACAGTGACAATAAT---GCCGGCTTAAC 3660  
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148  
 Db 3661 ATCGATGCAAAAAATGTAACAGTAACAACAATAATTACTTCTCACAAGCAGTGGAGCATC 3720  
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168  
 Db 3721 -----TCTGCCGACAGTGGAGAAATATACCACATAAACAAGGTACACCAATT 3765  
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188  
 Db 3766 AACGCAACCACTGGTAACGTGGAGATACCTCCCTCAACAGGTTAGT-----ATC 3813  
 QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203  
 Db 3814 CTAGGTGGAATTTAGTCCAGCTCTGGCTCTCTAACAACCTTACTGCAACGCGGGCTCTT 3873  
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223  
 Db 3874 GCTGTAAACAATATTTCC-----GGCAACACC 3900  
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243  
 Db 3901 GTTACTGTACTTGCATAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAATAATA 3957  
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263  
 Db 3958 GGAACCGGAGGTGAACCACTTCAAGTCAATCAGCGGATATCGCGGTACCAATTTCTGGT 4017  
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280  
 Db 4018 GGCAGTAGAGGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAAATTCAAAATAATA 4077  
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300  
 Db 4078 GCAACCAACAGCGGAGGTAAAGTAAACAGGTGCAACAGGTACAAATTTGGTGTACGATTTC 4137  
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320  
 Db 4138 GGTAAATCGGTAAAT-----GTTACGCGCAACCGCTGC 4170  
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340  
 Db 4171 GATTTAAACAGTTGGGAATGGCGCAGAA-----ATTAATGCGACAGAA 4212  
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360



```

QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Db 4078 GCAACAACAGCGAGCGTAAACGTAACAAGTGCACACAGGTGATGTTGGTACGATTTC 4137
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLysGln 320
Db 4138 GGTATACGTTAAAT-----GTTACGCAAAAGCTGGC 4170
QY 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 4171 GATTTAACAGTGGCAATGGCGCAGAA-----ATTAATGCCACAGAA 4212
QY 341 SerAlaAlaAlaGlyThrAspGlyLysLeuProSerMetMetProLleGlnThr 360
Db 4213 GGAGTGCACCTTAACATACATCATCGGCGCAATTA-----ACT 4251
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
Db 4252 ACCGAAGCTAGTTACACATACATTCAGCCAGGTCAGGTAATCTTCAGCTCAGGAT 4311
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrThrSerGluGlyLysLeu 398
Db 4312 GGTAGCGTTGCGAGGAATTAATATGCCCAATGTGACACTAAATACTACAGGCACCTTA 4371
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4372 ACTACCGTGAAGGTTCAACATTAATCAACACGCGTACCTGGTTATTATTAACGCAAAA 4431
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
Db 4432 CACGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGCTGGTAAATCAACCAACGCA 4491
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4492 AATGCTCCGCGAGCGTAAATCGACACACCTCAAG-----AGAGTGAACATCACT 4542
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 4543 GGG-----GATTTAATCACAATAATGGAATTA-----AATATC---ATT 4578
QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
Db 4579 TCATAAAGCGTATAAACACCGTACTGTTA---AAAGCGGTAAATATGATGTAATAC 4635
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 4636 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTGAAGCAACGC 4683

```

## RESULT 22

```

US-09-453-702B-39
; Sequence 39, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:

```

```

; APPLICANT: Blattner, Frederick R.
; Burtner, Valerie
; Perna, Nicole T.
; Flunkett, Guy
; Welch, Rod

```

```

; TITLE OF INVENTION: No. 6365723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Quarles & Brady
; STREET: 1 South Finckney Street
; CITY: Madison
; STATE: WI

```

```

; COUNTRY: US
; ZIP: 53701-2113

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 39:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 25165
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-453-702B-39

```

```

Alignment Scores:
Pred. No.: 8,27e-06 Length: 25165
Score: 165.00 Matches: 145
Percent Similarity: 33.70% Conservative: 71
Best Local Similarity: 22.62% Mismatches: 189
Query Match: 6.38% Indels: 236
DB: 3 Gaps: 35

```

```

US-10-009-823A-1 (1-502) x US-09-453-702B-39 (1-25165)

```

```

QY 8 GlyAlaThrGlyMetLysThrHis-----SerThrGlyLeuGlyThrValSer 23
Db 13018 GGAACACGCGGCGCTACCATCAGGTGACGTCATACCGGCTCCGACCATTACC 13077
QY 24 -----AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGln 37
Db 13078 TTTAAGCCATCAGCGCGCATACATCTGAACCGCGATGAAGGCG----- 13125
QY 38 GlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIle-----GlySerThrGly 55
Db 13126 -----CAGCGCTTGACCATCAGCGCGGCGAGTACGCGG 13158
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 13159 -----CTGCGCAGCGCGCGGCGGAGTTC-----ACGTCACG 13188
QY 76 ThrGlnGly-----AlaPheGluProGlyAsn----- 84
Db 13189 CTCAACGGTCAACAATACAGCGCCACACCGACGATCGGCGCAACTGGACCTTAACCGTG 13248
QY 85 SerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluAspLys 104
Db 13249 CCGGTGAGCGATTCGCGCGCATTTAGTCAGCCCACTATACGTCAGCGCCGCGCCAC 13308
QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124
Db 13309 -----AGTGCAGCAGGCAACACACCGCCAGCAGCGGCGAATTTACTGGTCGAC--- 13356
QY 125 SerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThr 144
Db 13357 -----AGCGCGCTGCGGCGGAGCTCAACATCAACAC 13386
QY 145 LeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThr 164
Db 13387 GTGGCA-----GGCGACGATATTATCAACCGCGCGGAGCGGGGCC 13428
QY 165 SerThrAlaLeuAsnAlaValValAsn-----LeuGlyAspSerThrAspLysThr 181
Db 13429 GATCAACCATCAGCGGGGTGGTGACTCGCGCGCGCGTGGCGATACCGTACCGTGACG 13488

```

QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrPro 201  
Db 13489 CTGGGC--GGGAACACTTACACCGCTACGTACAG-----13521  
QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln---220  
Db 13522 -----AGCACTTAAGCTGGAGCGTTCAGCGGTTCGACGCGGATCTCCAGCG 13569  
QY 221 ---GlyAsnSerHis---AspIleThrValTyrPheAspGlyAlaProSerSerThrGly 238  
Db 13570 TTGGGCAATGGTATTGACCAATTACCGCTCGGTCCACCAACGCTAATGCGAACACCGGG 13629  
QY 239 SerLysThrPheGluTyrLeuVal-----246  
Db 13630 AGCGCACGCGGATATCACCATTGATGCCAACCTGCCGGGCTGCGGTAGATACCGTG 13689  
QY 247 -----AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThr---Asp 261  
Db 13690 GCGGCGGATGATATCGTCAACAGCATCGACGCGGCGGTGGTGTATCACCAGCGGC 13749  
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281  
Db 13750 AGTAGCGGCTGAATGCAGGTGCTGTGCTGAGCGGTACCATCAACAGT-----13797  
QY 282 MetThrAlaPheThrProThr-----GlySerAlaThr 292  
Db 13798 ---GTGGGTATTCGCCACCGTGCAGCGGCGAGCGAGCTGGAGCGTTCGATTCGCGG 13854  
QY 293 LysAspLeuAsnAlaTyrGlnProAlaProLeu-----ValAsnGlyLeuProGln 309  
Db 13855 GCAACGCTCAGCGCTGCGTGGCGGCGGTTAACCGTGGAGGTAGACGGG-----CAA 13908  
QY 310 PheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThrLeuAspPhe-----326  
Db 13909 AGCAGCGGCAATAACCCAGTCAGCGTCAGCATCCGTCGATTTAACGGCGGTG 13968  
QY 326 -----326  
Db 13969 GCAATCAGCATCAACACCGTTGCCAGCGACGACGTGATTAAACGCGCGAGAAAGGCACC 14028  
QY 327 -----GlyIleLysSerGlnGlnAsnMet-----334  
Db 14029 AATCTGACTCTTTCGGCGAGTACACGCGGATTCAGAGCGGCGGCAACCGTCACT 14088  
QY 335 TrpAlaGlyAla-----ProAlaSerAlaAlaIleGlyThr-----347  
Db 14089 TTTGCGGTAAACCTTACACTGCAAGCTGCGCGCAACGCGAGCTGAGTGTAAACGTT 14148  
QY 348 -----AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364  
Db 14149 CCGCGCGGAGATCTGGCAACTCTGCCAGAGCGCGCGGCGAATGTGCGAGCGCGGTAGC 14208  
QY 365 SerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp-----380  
Db 14209 AGCGGAGCGGTAAACAGTGCCTCGCGACCCATCGGTATAGCGTTCAGCGCGCGCGCG 14268  
QY 381 -----GlyTyr 382  
Db 14269 AGCGTCAACATTACACCATCGCCAGCAGCATATCTTAAACGCGCGAGAGCGGAAGC 14328  
QY 383 Pro-----GlnGlyAspLeuValAspValThrIle 392  
Db 14329 CGGCTCAACATCAGCGGCGACGACGACCGCGCAACCGGCGGACGCGTACCGCTT 14388  
QY 393 -----ThrSerGluGlyLysLeuGln-----GlyLysTyrSerAsnSerGlnVal 407  
Db 14389 AACGGCGCAACCTACACCGGCACTGTGACGCGGACGCTAGCTGGAGCGTCAG-----14442  
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427  
Db 14443 -----GTTCCC-----ACTTCAGCCCTCGCGCGCTCAACGCA 14475

QY 428 GlyAsnAsnHistyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447  
Db 14476 AGCAATTACCGTCAGCGCACCGCTCAATGAC-----AAAGCGGCAACCGCGC 14526  
QY 448 ThrSerAsnTyrGly-----LysLeuSerValAsnGlnLeu 459  
Db 14527 AGCGGCGCATATCTGGCGGTAGACACACCGCGCGGTTCCTACCATTAACCGGTG 14586  
QY 460 GluThrSerAsnValAspMetSerArgLysMetValAsnMetIleIleGlnArg---478  
Db 14587 GCGGCGGATGATCATCATCAACGATGCGGCAACATGCGGAGCGGTGGTGTCTCCGGCAC 14646  
QY 479 -----GlyPheGlnMetAsnSerLysSerValThr 488  
Db 14647 AGTAGCGCGGGAAGCGGCGATGTGTGTGAGCGTGTGTCTCAACGCGCAACCTACAC 14706  
QY 489 Thr 489  
Db 14707 ACC 14709  
RESULT 33  
US-09-252-991A-13774/c  
; Sequence 13774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13774  
; LENGTH: 4188  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13774  
Alignment Scores:  
Pred. No.: 1.52e-06 Length: 4188  
Score: 160.00 Matches: 111  
Percent Similarity: 35.67% Conservative: 62  
Best Local Similarity: 22.89% Mismatches: 180  
Query Match: 6.13% Indels: 132  
DB: 4 Gaps: 27  
US-10-009-823A-1 (1-502) x US-09-252-991A-13774 (1-4188)  
QY 53 SerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal---71  
Db 1938 GCCACCGAGCGCTCCGGCAACACCCAGTGGCGGAGCAGTGTCCCGTGGACTCGGTAGCC 1879  
QY 72 -----ArgThrIlePheThrGlnGlyAlaPheGluPro 82  
Db 1878 CCGGCCACCGCAGTGATCAACCCCGACGACCGCACCGTCTACGCGGACCGCGGAGCG 1819  
QY 83 GlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGlu 102  
Db 1818 GCGACGACGCTGACCTCAGCGATGGCAACCGCACCGGATCGCGCAGGTCTCCGCGGAC 1759  
QY 103 AspLysValHistyrThrArgAlaGlyAsnPheArgPheThr-----Gln 117  
Db 1758 GGC-----AGCGCAACTGGAGCTTACCCCGTCCACCGCGTGGCG 1717  
QY 118 AspGlyPheLeu-----AsnAspProSerGlyPheThrLeuMetGlySer 132  
Db 1716 GATGGAACCGTGGTCAACGCGCACCGGATCGCGCGGCGCAACACC-----1669  
QY 133 ArgIleSerAsnAsnProAsnIleLysGluThrLeuGluProIleGlnLeuAspPhe 152



QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90  
Db 241 GAATATACGATTTTACCTAGCTTCTTTAAAGGCCACAAACAATCTTTGGATTGGCA 300  
QY 91 IleGlyGlyLysGlyPheGlnVal-----ThrLeuGluAspLysValHisTyrThr 108  
Db 301 ATGACTAGAGAAGATGCTTTTATTTGGTTTCAGACCAAGATGGAGAGTAAGATTAAAC 360  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 361 AAAGATGGAATTTTCACTTGATGATGAGGTTATTTGGTAAATAGCAAGGATACAAG 420  
QY 129 LeuMetGlySerArgIleSerAsnAsnPro 138  
Db 421 GTATTAGTAGTATTATTTTAAATAATCCT 450

RESULT 35  
US-08-483-857-2  
; Sequence 2, Application US/08483857  
; Patent No. 6020125  
; GENERAL INFORMATION:  
; APPLICANT: Chan, Voon Loong  
; APPLICANT: Louie, Helena  
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF  
; TITLE OF INVENTION: CAMPYLOBACTER  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/483,857  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-504  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 810 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-483-857-2

Alignment Scores:  
Pred. No.: 3,598-07 Length: 810  
Score: 155.00 Matches: 46  
Percent Similarity: 49.33% Conservative: 28  
Best Local Similarity: 30.67% Mismatches: 64  
Query Match: 5.99% Indels: 12  
DB: 3 Gaps: 5

US-10-009-823A-1 (1-502) x US-08-483-857-2 (1-810)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20  
Db 1 ATGCAAAATGATATTATCAAGCACTCGCGGAATGGTAACCTAGTTTAAATAACTTGAAT 60  
QY 21 ThrValSerAsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnValVal 40

Db 61 GTGATTACTATAATCTTGGCAATATCAATACAAGTGGATATAAAAGAGATGATGGTT 120  
QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53  
Db 121 ATTGCAGATTTTAAAGGATTTTAAAGAACTCAGGATGAGTTGCCTATAGAAAATCAC 180  
QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70  
Db 181 ACAAGAGATGATCTCGTTTGTAAATCTACTCAATAGATGGAATCCCAAGTTCTTCAA 240  
QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90  
Db 241 GAATATACGATTTTACCTAGCTTCTTTAAAGGCCACAAACAATCTTTGGATTGGCA 300  
QY 91 IleGlyGlyLysGlyPheGlnVal-----ThrLeuGluAspLysValHisTyrThr 108  
Db 301 ATGACTAGAGAAGATGCTTTTATTTGGTTTCAGACCAAGATGGAGAGTAAGATTAAAC 360  
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128  
Db 361 AAAGATGGAATTTTCACTTGATGATGAGGTTATTTGGTAAATAGCAAGGATACAAG 420  
QY 129 LeuMetGlySerArgIleSerAsnAsnPro 138  
Db 421 GTATTAGTAGTATTATTTTAAATAATCCT 450

RESULT 36  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Alignment Scores:  
Pred. No.: 1,11 Length: 4403765  
Score: 151.50 Matches: 119  
Percent Similarity: 31.12% Conservative: 64  
Best Local Similarity: 20.24% Mismatches: 220  
Query Match: 5.86% Indels: 185  
DB: 3 Gaps: 24

US-10-009-823A-1 (1-502) x US-09-103-840A-2 (1-4403765)

QY 3 GlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22  
Db 368448 GGCAATATCGCATCGGCTCACCGGC-----ACCGTTCAGATCGGGTTCGGCAGCTTC 368395  
QY 23 SerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGln 42  
Db 368394 AACTCGGCGGACCCACACATCGCTTGTTCACCTCGGTGACGGAAACCGTAGGATTCTTC 368335  
QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55  
Db 368334 AACTCGGCGACCCGCAACCGTGGGCATCGGAAACACCGGACCCGCAAACTTCGGCATCGCA 368275

QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75  
 DB 368274 AACTCGGCGGCTTCAACACCGGCTCGGG---AACACGGGAGACCAACACG--- 368224  
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys--- 94  
 DB 368223 ---GGCCTGTTCAACCGCGGCAACGTCACACCGGCTCGGCAACACCGGCGATC 368170  
 QY 95 ---GlyPhePheGlnVal--- 99  
 DB 368169 AACACGGGAGATCAACACCGGAGCTTCAACACTGGCAGCACCACATACCGGAGCTTC 368110  
 QY 100 ThrLeuGluAsp---LysValHisTyrThr 108  
 DB 368109 AACCTCGGCGATCAACACCGGCGAGTTCAACTCGGCTGACTACACACCGGCTACTTC 368050  
 QY 109 ArgAlaGlyAsnPheArgPhe--- 115  
 DB 368049 AACCGGCGTGACTACACACCGGCTGTGCCAACACCGGCAACGTCACACCGGCGCTTC 367990  
 QY 116 ---ThrGlnAspGlyPhe--- 120  
 DB 367989 ATCTCGGCAATTACAGCAACGGGTTCTTCTGGCGAGTGACTACCGGGGTGATTGGC 367930  
 QY 121 ---LeuAsnAspProSer 125  
 DB 367929 CTTTCCACACGATCACCAATTCGGAATCCCTACCGCTACGACTTGAGTGTTCACATC 367870  
 QY 126 GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsn--- 139  
 DB 367869 GACATACCCATCACCGGCGACGCTGTCGCCACCGCAACAGTTTCAACATTCACCATTC 367810  
 QY 140 ---IleLysLysGluThrLeuGluPro 147  
 DB 367809 TTCCAGATACGAGTCTTGCTGCTCGCGGGTGTGTCACAGAGATGATCGGCGCC 367750  
 QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167  
 DB 367749 ATCAGATCGATGTCATCAAGTCATCGCATCGCATCGCATCGCATCGCATCGCATCGC 367690  
 QY 168 LeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187  
 DB 367689 ATGGTTGGCACCGCGGCTTCGCGCGATCCCATCGCATCGCATCGCATCGCATCGC 367630  
 QY 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207  
 DB 367629 GGTTC---GGCAACTCGAC--- 367612  
 QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227  
 DB 367611 ---ACGCGCGCTGCTCGGTTCTTCCACACCGCGCGCGCATGATCGGC 367561  
 QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247  
 DB 367560 TTCGGGAATTCGCGCGCGCAACATGTCGGGCTCGCGGAATTC--- 367516  
 QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGly---ThrAspSerAlaGlyLeuLeu 266  
 DB 367515 ---GGCGTGGCAATTCGGGCTCTTTAACCGCGCGGCTTGGC 367474  
 QY 267 MetSerGlyThrMetThrPheSerSer---AsnGlyGluLeuLysAsnMetThrAlaPhe 285  
 DB 367473 AATTCGGGCTTACTGAATTCGCGCGCTGCGAGTCTGGCGAATTCGGCAACCTGGCAACAC 367414  
 QY 286 ThrProThrGlySerAlaThrLysAlaLeuAsnAlaTrpGlnProAlaProLeuValAsn 305  
 DB 367413 ATCTGGGCGTCTACACACGACGACGCTGACCTCGCGCGGCTGTTCTCTCGACACACCGGCAACCTG 367357  
 QY 306 GlyLeuProGlnPheSerAlaAsnPheValGlyValGlyIleGln---ProLeu 322  
 DB 367356 GGCAATCGAAACATCGCGCGCAACCTGCGCGGCTGTTCTCTCGACACACCGGCAACCTG 367297

QY 323 ThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342  
 DB 367296 ACCTGAGACTTCGGCTGCCAACACGAGCGGCTCAACCGCGGC--- 367252  
 QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362  
 DB 367251 ---ATCGGAACCTTGGCGAGCTCAACATCGGCTTCGTTAATACC 367210  
 QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgTyr--- 377  
 DB 367209 GGGACTCCAACTGGGCACTCGGCAACCTCGGCGACCTCACTTCGGGGGTCAACATC 367150  
 QY 378 ---SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395  
 DB 367149 GCGGGTAACAAACATCGGCTGCCAACACCGGATCTTCGATATCGGCTTGGCGAACCTG 367090  
 QY 396 GlyLysLeuGlnGlyLysTyrSerAsn---SerGlnValValAspPhe--- 410  
 DB 367089 GGCAGCTACAACTCGGTTGGCAATCTGGCGCAGCAACCTGGGCTTTGGCAACGCC 367030  
 QY 411 ---TyrAsnIleProLeuAlaArgPheThrSerGluAsp--- 422  
 DB 367029 GGCAGCTACAACTCGGCTTCGGAACCTCGGCGAGCAACCTGGGCTTTGGCAACACC 366970  
 QY 423 ---GlyLeuArgArgGluGlyAsnAsnHisTyrSer--- 433  
 DB 366969 GGCAGCTACAACTCGGCTTCGGAATACCGGTAAACAAACATCGGCTCGGCTCACC 366910  
 QY 434 ---AlaThrLeuAspSerGlyGlyProGluPheGlyLeu--- 445  
 DB 366909 GGCACCGCCAGATCGGATCGGAGCTCACTCGGCGAGCAACATCGGCTGTTTC 366850  
 QY 446 ---ProGlyThrSerAsnTyrGly 452  
 DB 366849 AACTCGGCGAGGAAACATCGG 366826  
 RESULT 37  
 US-09-103-840A-1/c  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37Rv  
 US-09-103-840A-1  
 Alignment Scores:  
 Pred. No.: 1.11 Length: 4411529  
 Score: 151.50 Matches: 119  
 Percent Similarity: 31.12% Conservative: 64  
 Best Local Similarity: 20.24% Mismatches: 220  
 Query Match: 5.86% Indels: 185  
 DB: 3 Gaps: 24  
 US-10-009-823A-1 (1-502) x US-09-103-840A-1 (1-4411529)  
 QY 3 GlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22  
 DB 368391 GGCAATATCGGATCGGCTCACCAGC-----ACCGTCAAGATCGGCTTCGCGAGCTTC 368338

QY 23 SerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnInValValPheGln 42  
Db 368337 AACTCGGAGCAGCCACCATCGCTTGTAACTCCGGGTGACGGAACAGTAGATTCTTC 368278  
QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55  
Db 368277 AACTCGGAGCAGCCACCATCGCTTGTAACTCCGGGTGACGGAACAGTAGATTCTTC 368218  
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75  
Db 368217 AACTCGGAGCAGCCACCATCGCTTGTAACTCCGGGTGACGGAACAGTAGATTCTTC 368167  
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys--- 94  
Db 368166 -----GGCCTGTTCAACCGCGGCAACGTCACACCGCGCGTCGCGCAACACCGCGCAGCATC 368113  
QY 95 -----GlyPheGlnVal----- 99  
Db 368112 AACACGGGAGCATCAACACCGCGAGCTTCAACACTGGCAGCACCAATACCGGAGCTTC 368053  
QY 100 ThrLeuGluAsp-----LysValHisTyrThr 108  
Db 368052 AACTCGGAGCAGCCACCATCGCTTGTAACTCCGGGTGACTACAAACACCGGCTACTTC 367993  
QY 109 ArgAlaGlyAsnPheArgPhe----- 115  
Db 367992 AACCGGGGTGACTACAAACCGGCGTGTGCGCAACAGCGGCAACGTCACACCGCGCGCTTC 367933  
QY 116 -----ThrGlnAspGlyPhe----- 120  
Db 367932 ATCTCCGGAATTACAGCAACCGGTTCTTCTGCGGAGGTGACTACACAGGGGTTGATTGGC 367873  
QY 121 -----LeuAsnAspProSer 125  
Db 367872 CTTTCCACACAGATCACCATTCCCGAATCCCTTACCGCTACGACTGAGTGTTCAAATC 367813  
QY 126 GlyPheThrLeuMetGlySerArgIleSerAsnProAsn----- 139  
Db 367812 GACATACCATCACCAGCAGCGTGTGCGCACCAACAGTTCACCATTCGCCAGC 367753  
QY 140 -----IleLysLysGluThrLeuGluPro 147  
Db 367752 TTCCAGATACAGTCTTCTGTCCTGCGCGGTGCTTGTCAACAGAGATGATCGGCCCC 367693  
QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167  
Db 367692 ATCAGATCGATGTCAATCAATGATCGCATCGCATCGCTTCCGCAACACCATCAGC 367633  
QY 168 LeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187  
Db 367632 ATGGTTGGACCGCGCGCTTCCGCGCCGATCCCATCGGCATCAGCATCGGTGTACCCCG 367573  
QY 188 TyrPheAlaLeuLeuGluSerTriLysGlyAsnGlyThrProPheSerThrSerAsn 207  
Db 367572 GGTTC-----GGCAACTCGACC----- 367555  
QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227  
Db 367554 -----ACGCGCGCTGCTCGGGTTCTTCCACACCGCGCGCCCATGTATCGGGC 367504  
QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247  
Db 367503 TTCCGGAACCTCGCGCGCGCAACATGTCGGGCTCGCGAATCTC----- 367459  
QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGly-----ThrAspSerAlaGlyLeuLeu 266  
Db 367458 -----GGCGCTGGCAATTCGGGCTTCTTTTAAACCGCGCGCTTGGGC 367417  
QY 267 MetSerGlyThrMetThrPheSerSer-----AsnGlyGluLeuLysAsnMetThrAlaPhe 285  
Db 367416 AATTCGGGCTTACTGAATTTCGGCGCGTGTGCGAGTCGGGTCTGGCGAACCTGGCGACAC 367357  
QY 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305

Db 367356 ATCTCGGCGGTCTCAACACAGCAGCAGCTGACCTCCGAGCGCCGCC-----TTCCGGCTCG 367300  
QY 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln-----ProLeu 322  
Db 367299 GGCATCGCAACATCGCGCGCAACCTCGCGCGCTGTTCCTCGACACACCGCAACCTG 367240  
QY 323 ThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342  
Db 367239 AGCTGAACTTCGCGGTGCGCAACAGCGCGGCTCAACCGCGGC----- 367195  
QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362  
Db 367194 -----ATCGGAACCTCGCGAGCTCAACATCGGCTTCGTTAATACC 367153  
QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyr----- 377  
Db 367152 GCGGACTCAACCTGGGATCGGCAACCTCGCGACCTCAACTTCGCGGGGTCAACATC 367093  
QY 378 -----SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395  
Db 367092 GCGGTAAACATCGGCGCATCGCCACACACCGGATCTTCGATATCGGCTTGGCGAACCTG 367033  
QY 396 GlyLysLeuGlnGlyLysTyrSerAsn-----SerGlnValValAspPhe----- 410  
Db 367032 GCGACTCAACATCGGCTTGGCAATCTCGCGACGACAACTTCGGCTTGGCAAGCC 366973  
QY 411 -----TyrAsnIleProLeuAlaArgPheThrSerGluAsp----- 422  
Db 366972 GCGAGTACAACTCGGCTTCGCAACTTCGCGAGCGCAACACTTCGGCTTGGCAACACC 366913  
QY 423 -----GlyLeuArgArgGluGlyAsnAsnHisTyrSer----- 433  
Db 366912 GCGAGTCAACATCGGCTTCGCAATCTCGGATACCGTAAACAAACATCGCGCTCGGGCTCACC 366853  
QY 434 -----AlaThrLeuAspSerGlyGlyProGluPheGlyLeu--- 445  
Db 366852 GCGAACGCCAGATCGGATCGGCGCTCAACTCGGCGAGCAACACATCGCGCTGTTTC 366793  
QY 446 ---ProGlyThrSerAsnTyrGly 452  
Db 366792 AACTCGCGCAGCGGAACATCGGG 366769

## RESULT 38

US-09-206-942-64  
; Sequence 64, Application US/09206942  
; Patent No. 6432569  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 3543  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-64

## Alignment Scores:

Pred. No.: 1.57e-05 Length: 3543  
Score: 149.50 Matches: 113  
Percent Similarity: 37.09% Conservative: 91  
Best Local Similarity: 20.55% Mismatches: 220  
Query Match: 5.78% Indels: 127  
Gaps: 21

```

US-10-009-823A-1 (1-502) x US-09-206-942-64 (1-3543)
QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerClnAspLeu 49
DB 1816 AATATTTCAGGCTTCACAAAGCAAGATTGTAGCTAAGAT-----AGTAGTAATTTA 1869
QY 50 AlalleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
DB 1870 ACTATTGGTAATAGTGATGATAGCGCAATACT-----1902
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
DB 1903 ACGCTAAACAGTA-----ACTTTTAACAATGTGTTAAAGATTCA 1941
QY 90 AlalleGlyGlyLeGlyPheGlnValThrLeuGluAspLysValHisThrArg 109
DB 1942 AAAATCTCTGCTGACGGT--CACAAAGGTGACACTAAATAGCAAGTGAAA--ACACTT 1995
QY 110 AlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeu 129
DB 1996 AGTGATAATGATAACAAACACTGAAGTGGCGAGTGACACAAAT---ACGGTGTAACTATT 2052
QY 130 MetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGln 149
DB 2053 ACTGCAAAAGATGTAGAGTAAACACAAATATTACT-----2088
QY 150 LeuAspPheAsnAspProThrValAlaLysSerProAla-----Lys 163
DB 2089 -----TCTCAAAACAGTGAACGCTCTCTCGGCAAAATGAGGAGTATACCACATAA 2139
QY 164 ThrSerThrAlaLeuAsnAlaVal-----ValAsnLeuGlyAspSerThrAspLys 180
DB 2140 ACAGGTACACCAATTAATGCAACCGCGGTAACTGGAGAGTAAACCGCTCATACAGGCAGT 2199
QY 181 ThrGln-SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGly-----197
DB 2200 ATCCAAGCGGAATTGAGTCAAGCTGGCTCTGTGACAAATGTTGGCAGCGCGGATACT 2259
QY 198 -----AsnGlyThrProIleSerThrSerAsn-----207
DB 2260 CTTCGTGTAGTAATATTTCAGGCAACCGCGGTACTCTT-ACGTGCAAAATAGCGGTGCATT 2318
QY 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnG 220
DB 2319 AACCACTTTGCGAGGCTCTACATTAAGGACCGAGAGTAAACCACTTCAAGTCAATC 2378
QY 220 nGlyAsn-----SerHisAsp 226
DB 2379 AGGTAATATCGCGGTAAAAATTTCCGGCAAGACAGATAAACGTTAAAGCACTAATAGTTT 2438
QY 226 eThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVa 246
DB 2439 AACCAACCAAGCAGACTCAAAATTTAGAGCATCTGAGGC-----GAGCTAATGT 2489
QY 246 lAlaMetAsnProSerGluAspGlySerAlaAlaSerGly-----259
DB 2490 AACAAAGCAAAACAGCAATAATTTGGCGGTCAATTTCTGTGGCACAGTAGTAGTACCAG 2549
QY 260 -----ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSe 275
DB 2550 GACCGAAGTTTAAACCCACAGAGGCTCTACGATTACTGGAAACGAGAGCGGTGACCA 2609
QY 275 rAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly-----SerAl 291
DB 2610 TTCAAGCAATCAGGTAATATCGCGGCGCATGATTCTGTGGCAAGTAGTAGTATAGCGC 2669
QY 291 aThrLysAspLeu-----As 296
DB 2670 AACCAAGATTTTAACTAAATCCGGTTCAGAGATTAAGCAACGCGCGGAGGTGAA 2729
QY 296 nAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValG 316

```

---

```

2730 TGTAAACAAGTGCACACAGGTACAAATTGACGGTACGATTTCCGGTAATACGGTAATGTTAC 2789
QY 316 YAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAl 336
DB 2790 AGCAAAATACTGGCGATTTAACTGTTGAAGATGCGCAAAA-----AT 2831
QY 336 aGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMe 356
DB 2832 TGATGCGACAGGAGGACCGGACCTTAATCTGCAACATCGGCAAAATTA-----2880
QY 356 tProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgAr 376
DB 2881 -----ACCACCTAAGCTAGTTTCAAGCATTAATCTCAGCTAATAACACAGGTAACCT 2930
QY 376 gTyrSerGlnAspGlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSe 394
DB 2931 TTCAGCTAAGGATGTAGCATTTGGGGGAATATCAATGCTGTAATGTAACACTGAATAC 2990
QY 394 rGluGlyLysLeuGlnGlyLysTy-SerAsnSerGlnValValAspPheTyrAsnIlePr 414
DB 2991 TACAGCGGCTCTAACTTACCGTGAAGGTTCAAGCATTAACGCCAACACAGCGGACCTTGGT 3050
QY 414 oLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAl 434
DB 3051 TATTAACGCAAAAGACGCTGAGCTAAATGGTGAGGCATCAGGTAACCACTAGTAGTAA 3110
QY 434 aThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLe 454
DB 3111 TGCAACCAACGCAATGCTCCGCGCAGCGTAACTCCGCAACACTCAACAGAGTCAACAT 3170
QY 454 uSerValAsnGlnLeuGluThrSerAsnValAsp---MetSerArgGluMetValAsnMe 473
DB 3171 CACTGGGATTTAATACATAAATGGATTAAATATCATTTTCAAAAAACGGTATATAACAC 3230
QY 473 tIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMe 493
DB 3231 CGTACTGTITA---AAAGCGGTAAATTAATGATGTGAATACATTCAACCGGTTAGCAAG 3287
QY 493 tLeuGlnLysAlaLeuGluLysArg 502
DB 3288 CGTAGATGAAGTAATTAAGCGAAACGC 3315

```

RESULT 39  
US-09-206-942-62  
; Sequence 62, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 62  
; LENGTH: 3568  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-62

Alignment Scores:  
Pred. No.: 1.59e-05 Length: 3568  
Score: 149.50 Matches: 113  
Percent Similarity: 37.09% Conservative: 91  
Best Local Similarity: 20.55% Mismatches: 220  
Query Match: 5.78% Indels: 127  
DB: 4 Gaps: 21

US-10-009-823A-1 (1-502) x US-09-206-942-62 (1-3568)

US-10-009-823A-1 (1-502) x US-09-206-942-62 (1-3568)

|    |      |                                                                      |
|----|------|----------------------------------------------------------------------|
| Qy | 30   | AsnThrIleGlyTyrLysGlnGlnInValValPheGlnAspLeuPheSerGlnAspLeu 49       |
| Db | 1841 | AAATATTTCAGGCTTCAACAAAGCAAAGATTGTAGCTAAAGAT-----AGTAGTAATTTA 1894    |
| Qy | 50   | AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69      |
| Db | 1895 | ACTATTGGTAATAGTAGTATACGGCAATACT-----1927                             |
| Qy | 70   | SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89      |
| Db | 1928 | AGCGCTAAACACGTA-----ACTTTTAAACAATGTTAAAGATTCA 1966                   |
| Qy | 90   | AlaIleGlyLysGlyPheGlnValThrLeuGluAspLysValHisThrArg 109              |
| Db | 1967 | AAATTTCTCTGCTGACCGT---CACAAAGTGACACTAAATAGCAAGTGAAA---ACRATT 2020    |
| Qy | 110  | AlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeu 129     |
| Db | 2021 | AGTGATNATGATAACAACACTGAAGGTGGCAGTGACACAAT---ACCGTTTTAACTATT 2077     |
| Qy | 130  | MetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGln 149     |
| Db | 2078 | ACTGCAAAAGATGTAGAAGTAAACACAATATTACT-----2113                         |
| Qy | 150  | LeuAspPheAsnAspProThrValAlaLysSerProAla-----Lys 163                  |
| Db | 2114 | -----TCTCAAAACAGTGAACGCTCTTCGGCAATGGAGGGATTACCACTAAA 2164            |
| Qy | 164  | ThrSerThrAlaLeuAsnAlaVal-----ValAsnLeuGlyAspSerThrAspLys 180         |
| Db | 2165 | ACAGGTACAACCACTAATATGCAACCGCCGGTAAACGCTGGAGATAACCGCTCATACAGCAGT 2224 |
| Qy | 181  | ThrGln-SerGluAlaAsnProTyrPheAlaLeuGluSerTrpLysGly 197                |
| Db | 2225 | ATCCAAGCGCGAATGTAGTCAAGCTGGCTCTGTGACAAATTGGCGCGCGCGGTACT 2284        |
| Qy | 198  | -----AsnGlyThrProIleSerThrSerAsn-----207                             |
| Db | 2285 | CTTGCTGTAGTAATATTTCAAGCAACGCGGTACTGTT-ACTGCAATAGCGGTGCATT 2343       |
| Qy | 208  | -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGly 220                     |
| Db | 2344 | AACCACTTTGGCAGGCTCTACAATTAAGGAACCGCAGAGTATACCACTTCAAGTCAATC 2403     |
| Qy | 220  | nGlyAsn-----SerHisAspIle 226                                         |
| Db | 2404 | AGGTAATATCGCGGTAAATTTCCGGCAACACACAGTAAACGTTAAACCACTAATAGTTT 2463     |
| Qy | 226  | eThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVa 246     |
| Db | 2464 | AACCAACCAACGACACTCMAAAATGAAGCGCACTGAAGGC-----GAGGCTAATGT 2514        |
| Qy | 246  | IalaMetAsnProSerGluAspGlySerAlaAlaSerGly-----259                     |
| Db | 2515 | AACAAGCAAAACAGACATAATTGGCGGTACATAATTTCTGGTGGCACAGTAGAAGTTACCGC 2574  |
| Qy | 260  | -----ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSe 275             |
| Db | 2575 | GACCGAAGTTTAAACCAACCAAGCAGGCTCTACGATTACTGGAACCGAGAGCGTGACCA 2634     |
| Qy | 275  | rAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly-----SerAl 291            |
| Db | 2635 | TTCAAGCCAATCAGGTAATATCGCGGCATGATTCTTGTTGGCAAGTAGAAGTTAGCGC 2694      |
| Qy | 291  | aThrLysAspLeu-----AS 296                                             |
| Db | 2695 | AACCAAGATTTAATCTAAATCCGGTTCAGAGATTAAACCAACGCGCGCGAGGTCAA 2754        |
| Qy | 296  | nAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGly 316    |
| Db | 2755 | TGTAACAAGTGCACAGGTACAATATGACGGTAGCAATTTCCGGTATACGTTAAATGTTAC 2814    |

|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| Qy | 316  | YAlAGlYlLeGlnPrLeuThrLeuAspPheGlylLeLysSerGlnGlnAsnMetTrpAl     | 336  |
| Db | 2815 | AGCAAACTACGGCGATTAACTGTTGAAGATCCGCCAAA                          | 2856 |
| Qy | 336  | aGlyAlaProAlaSerAlaAlaAlaLeGlyThrAspLeGlyLysLeuProSerMetMe      | 356  |
| Db | 2857 | TGATGCGACAGGAGGAGCGGACCCCTAACTGCAACATCGGCAAAATTA                | 2905 |
| Qy | 356  | tProLeGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgAr        | 376  |
| Db | 2906 | -----ACCATAAGCTAGTTTCAGCATTTACTTCAGCTAATAACCGAGTAAACCT          | 2955 |
| Qy | 376  | gTySerGlnAspGlyTyProGlnGlyAspLeu-----ValAspValThrIleThrSe       | 394  |
| Db | 2956 | TTCAAGTAAAGTAGGTAGCATTCGGGGAATATCAATGCTTAATGTAACATCGAATAC       | 3015 |
| Qy | 394  | rGluGlyLysLeuGlnGlyLysTySerAsnSerGlnValValAspPheTyAsnIlePr      | 414  |
| Db | 3016 | TACAGGCGCTCTAACTACCGTGAAGGTTCAAGCATTAACGCCAAACGGCGCACCTTGGT     | 3075 |
| Qy | 414  | oLeuAlaArgPheThrSerGlnAspGlyLeuArgArgGluGlyAsnAsnHisTySerAl     | 434  |
| Db | 3076 | TATTAAACGCAAAAGACGCTGAGCTAAATGGTGAGGCATCAGGTAAACATACAGTAGTGA    | 3135 |
| Qy | 434  | aThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyTyGlyLysLe   | 454  |
| Db | 3136 | TGCAACCAACGCAAAATGGCTCCGCGACGCGTAATCGGCACCAACCTCAAGCAGAGTGAACAT | 3195 |
| Qy | 454  | uSerValAsnGlnLeuCluThrSerAsnValAsp--MetSerArgGluMetValAsnMe     | 473  |
| Db | 3196 | CACTCGGGGATTTAATCAATAATGGATTTAAATATCATTTTCAAAAACGGTATAAACAC     | 3255 |
| Qy | 473  | tIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMe    | 493  |
| Db | 3256 | CGTACTGTTA--AAAGCGCTAAAAATTGATGTGAAATACATTCAACCGGGTATAGCAAG     | 3312 |
| Qy | 493  | tLeuGlnLysAlaLeuGlnLeuLysArg                                    | 502  |
| Db | 3313 | CGTAGATGAAGTAAATTGAAGCGAAACGC                                   | 3340 |

RESULT 40

```

US-09-252-991A-2729/c
; Sequence 2729 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2729
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-2729

```

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Alignment Scores:      | 6.06e-06 | Length:       | 1384 |
| Prod. No.:             | 145.50   | Matches:      | 103  |
| Score:                 | 34.63%   | Conservative: | 57   |
| Percent Similarity:    | 22.93%   | Mismatches:   | 159  |
| Best Local Similarity: | 22.93%   | Indels:       | 154  |
| Query Match:           | 5.63%    | Gaps:         | 18   |
| DB:                    | 4        |               |      |
|                        |          |               | DB:  |

US-10-009-823A-1 (1-502) X US-09-252-991A-2729 (1-1284)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:09:07 ; Search time 132 Seconds  
(without alignments)

1231.264 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLFIGATGKMTHTSTGLG.....NGSKSVTTADTLQKALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description      |
|------------|-------|-------------|--------|----|------------------|
| 1          | 2586  | 100.0       | 502    | 14 | US-10-009-823A-1 |
| 2          | 529.5 | 20.5        | 414    | 15 | US-10-009-823A-1 |
| 3          | 527.5 | 20.4        | 462    | 14 | US-10-009-823A-1 |
| 4          | 506.5 | 19.6        | 413    | 15 | US-10-009-823A-1 |
| 5          | 485   | 18.8        | 718    | 15 | US-10-009-823A-1 |
| 6          | 355.5 | 13.7        | 477    | 14 | US-10-009-823A-1 |
| 7          | 351   | 13.6        | 462    | 14 | US-10-009-823A-1 |
| 8          | 337.5 | 13.1        | 400    | 14 | US-10-009-823A-1 |
| 9          | 333   | 12.9        | 401    | 14 | US-10-009-823A-1 |
| 10         | 296   | 11.4        | 221    | 15 | US-10-009-823A-1 |
| 11         | 288   | 11.1        | 262    | 15 | US-10-009-823A-1 |
| 12         | 288   | 11.1        | 271    | 15 | US-10-009-823A-1 |
| 13         | 288   | 11.1        | 279    | 15 | US-10-009-823A-1 |

|    |       |      |      |    |                    |                   |
|----|-------|------|------|----|--------------------|-------------------|
| 14 | 282   | 10.9 | 480  | 14 | US-10-009-823A-11  | Sequence 11, Appl |
| 15 | 233   | 9.0  | 360  | 14 | US-10-009-823A-12  | Sequence 12, Appl |
| 16 | 201.5 | 7.8  | 152  | 15 | US-10-009-823A-13  | Sequence 13, Appl |
| 17 | 189.5 | 7.3  | 464  | 15 | US-10-009-823A-14  | Sequence 14, Appl |
| 18 | 186   | 7.2  | 370  | 14 | US-10-009-823A-15  | Sequence 15, Appl |
| 19 | 180.5 | 7.0  | 655  | 14 | US-10-009-823A-16  | Sequence 16, Appl |
| 20 | 180.5 | 7.0  | 655  | 14 | US-10-009-823A-17  | Sequence 17, Appl |
| 21 | 173.5 | 6.7  | 1095 | 14 | US-10-009-823A-18  | Sequence 18, Appl |
| 22 | 173.5 | 6.7  | 1095 | 14 | US-10-009-823A-19  | Sequence 19, Appl |
| 23 | 173.5 | 6.7  | 1095 | 14 | US-10-009-823A-20  | Sequence 20, Appl |
| 24 | 172.5 | 6.7  | 1536 | 14 | US-10-009-823A-21  | Sequence 21, Appl |
| 25 | 166   | 6.4  | 3705 | 15 | US-10-009-823A-22  | Sequence 22, Appl |
| 26 | 166   | 6.4  | 3705 | 15 | US-10-009-823A-23  | Sequence 23, Appl |
| 27 | 166   | 6.4  | 3705 | 15 | US-10-009-823A-24  | Sequence 24, Appl |
| 28 | 165.5 | 6.4  | 124  | 15 | US-10-009-823A-25  | Sequence 25, Appl |
| 29 | 164   | 6.3  | 343  | 15 | US-10-009-823A-26  | Sequence 26, Appl |
| 30 | 162   | 6.3  | 1954 | 15 | US-10-009-823A-27  | Sequence 27, Appl |
| 31 | 160   | 6.2  | 1230 | 9  | US-09-881-752A-150 | Sequence 28, Appl |
| 32 | 160   | 6.2  | 1230 | 9  | US-09-881-752A-150 | Sequence 29, Appl |
| 33 | 156.5 | 6.1  | 3073 | 15 | US-10-009-823A-28  | Sequence 30, Appl |
| 34 | 153.5 | 5.9  | 679  | 14 | US-10-009-823A-29  | Sequence 31, Appl |
| 35 | 151.5 | 5.9  | 269  | 15 | US-10-009-823A-30  | Sequence 32, Appl |
| 36 | 151.5 | 5.9  | 2204 | 15 | US-10-009-823A-31  | Sequence 33, Appl |
| 37 | 149.5 | 5.8  | 1180 | 14 | US-10-009-823A-32  | Sequence 34, Appl |
| 38 | 149.5 | 5.8  | 1188 | 14 | US-10-009-823A-33  | Sequence 35, Appl |
| 39 | 148.5 | 5.7  | 273  | 15 | US-10-009-823A-34  | Sequence 36, Appl |
| 40 | 147   | 5.7  | 1649 | 14 | US-10-009-823A-35  | Sequence 37, Appl |
| 41 | 146.5 | 5.7  | 486  | 14 | US-10-009-823A-36  | Sequence 38, Appl |
| 42 | 143   | 5.5  | 1331 | 15 | US-10-009-823A-37  | Sequence 39, Appl |
| 43 | 140.5 | 5.4  | 529  | 14 | US-10-009-823A-38  | Sequence 40, Appl |
| 44 | 139.5 | 5.4  | 1797 | 14 | US-10-009-823A-39  | Sequence 41, Appl |
| 45 | 139.5 | 5.4  | 1805 | 14 | US-10-009-823A-40  | Sequence 42, Appl |

ALIGNMENTS

RESULT 1

US-10-009-823A-1  
; Sequence 1, Application US/10009823A  
; Publication NO. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Ankenbauer, Robert G.  
; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES  
; FILE REFERENCE: DAVI150.001APC  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; TYPE: PRT  
; LENGTH: 502  
; ORGANISM: Lawsonia intracellularis  
US-10-009-823A-1

Query Match 100.0%; Score 2586; DB 14; Length 502;  
Best Local Similarity 100.0%; Pred. No. 1.6e-202; Indels 0; Gaps 0;  
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 MMGSLFIGATGKMTHTSTGLTGVNNIANANTIGYKQQVWFQDLFSQDLAIGTSGQPN 60  
DB 1 MMGSLFIGATGKMTHTSTGLTGVNNIANANTIGYKQQVWFQDLFSQDLAIGTSGQPN 60

```
QY 61 QAGMGAQVGSVRTIIFQGAPEFGNSVTDLAIGKGFQVLTLEDKVHYTRAGNFRFTQDGF 120
Db 61 QAGMGAQVGSVRTIIFQGAPEFGNSVTDLAIGKGFQVLTLEDKVHYTRAGNFRFTQDGF 120
QY 121 LNDPSGFTLMGSRISNNPNIKKTELPIQLDNDPDTVAKSAPKSTALNAVNLGDSXDK 180
Db 121 LNDPSGFTLMGSRISNNPNIKKTELPIQLDNDPDTVAKSAPKSTALNAVNLGDSXDK 180
QY 181 TQSEANPYFALLESKMGKNGTTPISSTNSYAPQMRVYDQGNSHDITVYFDGAPSSGSK 240
Db 181 TQSEANPYFALLESKMGKNGTTPISSTNSYAPQMRVYDQGNSHDITVYFDGAPSSGSK 240
QY 241 TFEYLIVAMPNPSDGSAAAGTSDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKOLNAWOP 300
Db 241 TFEYLIVAMPNPSDGSAAAGTSDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKOLNAWOP 300
QY 301 APLVNLGLOFNSANFVGAGIQLTLDFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQT 360
Db 301 APLVNLGLOFNSANFVGAGIQLTLDFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQT 360
QY 361 SSGNSTANGSSSTRYSODGYPQDGLVDVTITSEGLKQKYSNSQVVDVFNIPARPTS 420
Db 361 SSGNSTANGSSSTRYSODGYPQDGLVDVTITSEGLKQKYSNSQVVDVFNIPARPTS 420
QY 421 EDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETSNVDMREMNWIIIOQGF 480
Db 421 EDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETSNVDMREMNWIIIOQGF 480
QY 481 QNKSXSVTTADTLMQKALELKR 502
Db 481 QNKSXSVTTADTLMQKALELKR 502

RESULT 2
US-10-282-122A-48141
; Sequence 48141, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48141
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
; US-10-282-122A-48141

Query Match 20.5%; Score 529.5; DB 15; Length 414;
Best Local Similarity 29.8%; Pred. No. 1.4e-34;
Matches 149; Conservative 72; Mismatches 182; Indels 97; Gaps 13;

QY 8 GATGKTHGTGLGTSSNNIANNANTIGYKQQVQVFOOLFSDLAIGSTGSGGNQAGMAQ 67
Db 6 GLSGLAGASSNLDVIGNNIANNANTVGFQGRANFADMYANSVATSVN-----TGIGIGTR 60
QY 68 VGSVRTIIFQGAPEFGNSVTDLAIGKGFQVLTLEDKVHYTRAGNFRFTQDGLNDPDSGF 127
Db 61 LASVQNFQGGTINSTKSLDVAINGNGFFQSSNGVTYSRDTGTHRDKXGAIYDAQGR 120
QY 128 TLMGSRISNNPNIKKTELPIQLDNDPDTVAKSAPKSTALNAVNLGSDTKTQSEANP 187
Db 121 NLMGYAAGAGGVINTAQTVPLOA-----PTNNIAPRATS--KITQGFNL-NAQDKVPKAK--- 171
QY 188 YFALLESKMGKNGTTPISSTNS--YSYAPQMRVYDQGNSHDITVYFDGAPSSGSKTFEYL 245
Db 172 -----TPFNATDNTTNNSSLOVDTTGGSQQVTMYF--AKSAGT----- 211
QY 246 VAMPNPSDGSAAAGTSDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKOLNAWOPAPLVN 305
Db 212 -----WLAVAGVGGQPTTNLGTVTTFDASGRISSTTS--AATG-----QPTP--- 250
QY 306 GLPQPSANFV-----GAGTQPLTLDLFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQS 361
Db 251 SLGQFAFSIPTNTTGGANPQNLTLDL----- 275
QY 362 SGNSTANGSSSTRYSODGYPQDGLVDVTITSEGLKQKYSNSQVVDVFNIPARPTS 421
Db 276 -GGTQYGGKQGVNLAQDGFASGTLTTFISGTDGKLAGNSNGQSAVLGLIALANFNPN 334
QY 422 DGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETSNVDMREMNWIIIOQGFQ 481
Db 335 NGLVNIIGNQYAEATAASGVPOIAAPGSTHGTLOGSALENSNVNLTTLVNLITAQRYNQ 394
QY 482 MNSKSVTTADTLMQKALELK 501
Db 395 ANAQTIKQOAVDQTLNLNR 414

RESULT 3
US-10-127-032-101
; Sequence 101, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Banger, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; TITLE OF INVENTION: BIOFILM FORMATION
; FILE REFERENCE: UIZ-070CP
; CURRENT APPLICATION NUMBER: US/10/127,032
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,190
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/344,142
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 462
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-127-032-101

Query Match
Best Local Similarity 20.4%; Score 527.5; DB 14; Length 462;
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;

QY 4 SFLGATGKTHSTGLTVSNINANTIGYKQQQVFDLFSQDLAIGTSGSQPNQAG 63
Db :
2 SFNIGLSGIAASSGNLVTTGNINAGTVGFKQRAEFADVVAASV-LGS-----GSPQ 56
QY 64 MGAQGVSVATITQGAFFPGNSVTDLAIGCKGFFOVLTEDKVHYTRAGNFRFTQDGLND 123
Db :
57 SGVLLSDVSQMKQGNIDSTNSVLDLAINGFFVTSNNGAISYTRAGVENDKQDFIVD 116
QY 124 PSFTLMGSRISNNPNIKETLEPIQLD-FNDPTVAKSPAKTSTALNAVNLGDSDTKTQ 182
Db :
117 NNGYRLQGVAVGPNGLQNGVVTDLKVERANQAPQATSSIQOOSYNLNLTK-----P 168
QY 183 SEANPYFALLESKMGNGTPISTSNYSYAQPMRVYDQGNSHDITVYF---DGAPSS-- 237
Db :
169 PTVTPF-----DPSDAATYNSSSSGLIYDSQGNSHTMGSOFFIKNEPDPENATPP 216
QY 238 ---GSKTFYLV-AMNPSBDGSAAGTDSAGLLMSGTMTFSSNGELKNTAFTPTGS--- 290
Db :
217 IPENSWTMKVLIDGNPLDPSNKT-----MSFNVTFDASGQWTSVRA--PDGSTG 266
QY 291 -----ATKDLNAWQAPLVNGLPQPSANFVGAGIQLTLDFGIKSQONMWAGAPASAA 344
Db :
267 PGFSIDATTNVIQFSPA---TGNPPTP-----GNGWIPASD-GKPTPTVANGATGASG 318
QY 345 IGTIDIGKLPMSMPIQTSSGNSGNSTARNSSSTRYS-----QDGYPOGDLVDVTITSE 395
Db :
319 ISFDMRK-----TTQYSTAFQSNPIQDGYTTGQLAGLEIDT 356
QY 396 GKLGKYSNSQVDFNIPARTSBDGLRREGNNHYSATLDSGGPEFGLPGTSNVGKLS 455
Db :
357 GVIFARYTNGQSKVOGVVLNANFANIQGLTPIKTSVOSSESSEGEPAVCAPRSGTLGALQ 416
QY 456 VNOLETNSVDMSEMNMIIOQGFQWNSKSVTTADTMLOKALEK 501
Db :
417 SGALASNVDSINELNVLVHQRYQANAKTIQTEDAVTQTIINLR 462

RESULT 4
US-10-282-122A-50763
; Sequence 50763, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Orlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
```

```
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50763
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50763

Query Match
Best Local Similarity 19.6%; Score 506.5; DB 15; Length 413;
Matches 137; Conservative 81; Mismatches 180; Indels 105; Gaps 11;

QY 8 GATGKTHSTGLTVSNINANTIGYKQQQVFDLFSQDLAIGTSGSQPNQAGMAQ 67
Db :
6 GLSGLAGASSDLDVIGNNIANANTVGFKSTAQPSDMYANSVA-----SAVNNPIGIGTM 60
QY 68 VGSVRTITQGAFFPGNSVTDLAIGCKGFFOVLTEDKVHYTRAGNFRFTQDGLNDPSGF 127
Db :
61 LASVQQQFSSQITTSSTSLNVAINGNGFFQMSNNGVTVYSDGTFFQDKNGYIVNSQGL 120
QY 128 TLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNLGDSDTKTQSEANP 187
Db :
121 NLMGVAANANGVINTAATVPLQA---PTTNIAPTAT-TKITGQFNL-NSQDAYPA-TTP 173
QY 188 YFALLESKMGNGTPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSSTGSKTFEYLV 247
Db :
174 F-----NYTDPTS---YNYTTSVQVFDTLGSSQNNVLYFKVSATSGQWEAYAGPAG 221
QY 248 MNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNTAFTPTGSAATKDLNAWQAPLVNGL 307
Db :
222 KTPIDLGS-----VKFSTAGTITGTSTP----- 244
QY 308 PQFSANFVGAGIQLTLDFGIKSQONMWAGAPASAAIGTDIGKLPMSMPIQTSSGN--- 364
Db :
245 -----STARNGSSSTRYSQDGYPOGDLVDVTITSEGLQKYSNSQVDFYNIPLAR 270
QY 365 -----LTLDLTGTQYGGKNGINNLAQDGFASGVLTTFISIGADGKVTGNYSGQSTGLGQIVLAN 330
Db :
418 FTSEDGLRREGNNHYSATLDSGGPEFGLPGTSNVGKLSVNOLETNSVDMSEMNMIIO 477
Db :
331 FNNPGLVNVGNNQYVESAAASGVQISAPGSTNHTGLQSALENSNVLDTSOLVKLITAQ 390
QY 478 RGFQMSKSVTTADTMLOKALEL 500
Db :
391 RNYQANAQTIKQIVDQTLNL 413

RESULT 5
US-10-335-977-4820
; Sequence 4820, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
```

CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977  
 FILING DATE: 30-Dec-2002  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/993,002  
 FILING DATE: 17-DEC-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy E.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: GTN-018  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)742-7400  
 TELEFAX: (617)742-4214  
 INFORMATION FOR SEQ ID NO: 4820:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 718 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Helicobacter pylori  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (5) LOCATION 1...718  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4820:  
 US-10-335-977-4820

Query Match 18.8%; Score 485; DB 15; Length 718;  
 Best Local Similarity 25.0%; Pred. No. 1.3e-30;  
 Matches 181; Conservative 94; Mismatches 222; Indels 226; Gaps 22;  
 QY 1 MMSLFGATGKTHSTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGS----TGS 56  
 DB 1 MLRLWSGVNGMQAHLALDIESNNIANVNTGPKYSRASFDVMDLSQVKLIATAPYKNGL 60  
 QY 57 QGPN--QAGMGAQVGSVRTITQGAFFPGNSVTDLAIGGKGFQVLEDKV--HYTRAGN 112  
 DB 61 AGQDNFSGVLGVGDVATTKIFSQSQNTQNTDVKTDLAIGQDGFIIISPDRTGITEFTRDGE 120  
 QY 113 FRFTQDGLNDPSPGFTLM-----GSRISNNPNIKKETLEPIQLDFNDPPTVAKSPAK 163  
 DB 121 FLFDSQSLVTTGGLVGVQVRNGSDTGKSGDTDALKVDNTGPLENIRIDPGWV-WPAR 179  
 QY 164 TST--ALNAVNVLDGSDTKTQSEANFPFALLESWK--GNCTPPISTNSYSAQPMR---- 215  
 DB 180 ASNRISMRLNLRAGRHADQTA-----VFALDSSAKTPSDGINFVYDSGTNLACVAEDMGS 235  
 QY 216 VYDQGN-----SHDITVYFDGA-----PS-----STGSKTFEY 244  
 DB 236 LYNEGDGALLNENQGLWVYSKAKVVKDILPAENSTLELNGVKISFTNDSAVSRSTSL 295  
 QY 245 LVANWPEEDGSAAG-----TDSAGLLMSGTWTSSNGELKN----- 281  
 DB 296 VAAKNAINAVKSGTGIKAYLDGKQLRENTNELDGDEKLNIVVTQAGTGAFAFNLFDGDK 355  
 QY 282 -MTAF---TPTGSATKDLNAPAPLVNGLPQFSANFV-----GAGIQPLTL 324  
 DB 356 DVTAFKYSYTHSISPNADICQFRTEDRLALIGHDANIWKDPSLADNYQDSASIGVTIN 415  
 QY 325 DFG-----IKSQQNMW-----AGAPASAA 343

Db 416 QYGMFEINNKNKNIKENLNFVSGYSSDSVTNNVLFKNAMKGLNTASLEGGSASSS 475  
 QY 344 -----AIGT----- 347  
 Db 476 KPTHATHATSIDVSLGTGKHARIEFYRSGGAENFRVIVPEPGELVGGNAARNVEFG 535  
 QY 348 -----DIGKLPSMMP-----IQSSGNSTARNSSSTRYS-----Q 379  
 Db 536 GRILHFNNDGSLAGMNPPLQLQDFPKNGADAPORINLAFSGSGSFDGLTSDVKISYTAIEQ 595  
 QY 380 DGVPQGDLDVDTITSEGLQKGYNSQVDFYNIFLARFTSEDLRRRGNHYSATLDSG 439  
 Db 596 NGYQAGDLMDFRDSGVLGAFSNGRTIALAQVALANFANDAGLQAGNVFSQTNGSG 655  
 QY 440 GPEFGLPGTSNYGKLSVNQLETSNVDMGREMVNMIIIQGFQMSKSVTTADTMLQKALE 499  
 Db 656 QALIGAANTGRRGSGISGKLESSNVDSRLTNLIVQRFQANSKAVTTSQIILNTLN 715  
 QY 500 LXR 502  
 Db 716 LKQ 718  
 RESULT 6  
 US-10-009-823A-7  
 ; Sequence 7, Application US/10009823A  
 ; Publication NO. US20030157120A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Panaccio, Michael  
 ; APPLICANT: Rosey, Everett Lee  
 ; APPLICANT: Sinistaj, Meri  
 ; APPLICANT: Hasse, Detlef  
 ; APPLICANT: Parsons, Jim  
 ; APPLICANT: Ankenbauer, Robert G.  
 ; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE  
 ; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES  
 ; FILE REFERENCE: DAVI150.001APC  
 ; CURRENT APPLICATION NUMBER: US/10/009,823A  
 ; CURRENT FILING DATE: 2002-08-13  
 ; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
 ; PRIOR FILING DATE: 2000-05-11  
 ; PRIOR APPLICATION NUMBER: US 60/133,973  
 ; PRIOR FILING DATE: 1999-05-13  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 477  
 ; TYPE: PRT  
 ; ORGANISM: Treponema phagedenis  
 US-10-009-823A-7

Query Match 13.7%; Score 355.5; DB 14; Length 477;  
 Best Local Similarity 23.7%; Pred. No. 2.8e-20;  
 Matches 142; Conservative 68; Mismatches 171; Indels 217; Gaps 21;  
 QY 1 MMSLFGATGKTH-----STGLGTGVSNNIANAN 30  
 DB 1 MMSLFGVSQMNQHGVPKEVGLGVWVASGVNPKVEVGLGVWVASTPMVDVGNVANVN 60  
 QY 31 TIGYKQQ-----VVFDLFSQDLAIGSTGSGGNOAGMGAQVG 69  
 DB 61 TTGFRGRIDTVHTQGLQTTGINTDIAIVNFQLISQQL-----SGASRNE-----EVG 111  
 QY 70 SVRTITQGAPEPGNSVTDLAIGKGFQVLEDKVHYTRAGNFRFTQDGLNDP--SGFT 128  
 DB 112 Q-----GNQFFILKQGEKSPYTTAGAFGVDRDGTVLNPANGAC 149  
 QY 129 LMGSRISNNPNIKKETLEPIQ-LDFNDPPTVAKSPAKTSTALNA-----VNVLDGSTKQTS 183  
 DB 150 NLDKRLMRVQGMMAEDIEGQOIINTSDQPELPEGANQADILRSTEDLITPIGQKID---- 205  
 QY 184 EANPYFALLESWKNGTTPPISTNSYSAQPMRVVYDQGNSHDITVYFDGAPSGTG----- 238

Db 206 -----AKATTDVA---YTWATDENYDTFGQHKLQWVFPVPGNNQWLAT 249  
Qy 239 --SKTPEYLWAMPSPDSGSAAS-----GTDAG-LLMSGT-----MTFSSNGELKNMT 283  
Db 250 VTTDTAGNVTPNVDENQAGTTRVIGTITDAGQVLLVQATENTFIVFDNYGHLAS-- 307  
Qy 284 APTFTGSATKDLNAMPAPLVNGLPFSANFVGA-----GIQPIJTLDFGKSKQONWAGAP 339  
Db 308 -----SYNVGANPDGGAAPRHTTFNINDQSGIITGVY 340  
Qy 340 ASAAA---IGT-----DIGKLPSSMPEIQTSNGNSTARNSSSTRYSQDGYQP 384  
Db 341 SNGASLEGEIGTGRNITQPAEREIQOL-ALAGFANQGGLEKA--GESITIKAYQDGYAM 397  
Qy 385 GDLVDVITITSEGKLOGKYSNSQVVDYFNIPLARFTSEDLRREGNNHYSATLDSGGPEFG 444  
Db 398 GYLENFKIT-----YIQSNNSGIANIT 419  
Qy 445 LPGTSNYGKLSVNQLETSNVDMSREVMNMIIIOGFQMSKSVTTADTLMQKALELKR 502  
Db 420 VSGVMGKGKLIAGTLENSVNDLTDQFTDMIITQGFQAGAKTIQTSMTLTVLNLKR 477

RESULT 7

US-10-009-823A-8  
; Sequence 8, Application US/10009823A  
; Publication No. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Akenbauer, Robert G.  
; TITLE OF INVENTION: LAMSONIA DERIVED GENE AND RELATED FLGE  
; FILE REFERENCE: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 462  
; TYPE: PRT  
; ORGANISM: Treponema pallidum  
US-10-009-823A-8

Query Match 13.6%; Score 351; DB 14; Length 462;  
Best Local Similarity 23.0%; Pred. No. 6.2e-20;  
Matches 135; Conservative 67; Mismatches 176; Indels 208; Gaps 17;  
Qy 1 MMGSLFTGATGKTH-----STGLGTVSNNIANANTIGYKQOVVQDLF 45  
Db 1 MMRSLFSGVSMQNHQVNPKEVGLGVLIASTRMDVIGNNVANVTGPKR----- 51  
Qy 46 SQDLAIGSTGSGPNQAGMAQVSVRTIETQGAPEPNSVTLAI----- 91  
Db 52 -----GRDVTHTQALQTTGINTDVSIVNFDLISQOLSA 86  
Qy 92 -----GGKGFQVTLKDVHYTRAGNFRFTQDGLNDPDSGFT----- 128  
Db 89 ARPNEEVQSGGFFVLKSGEKTFTFRAGAGVDNAGTLVNPANGACNLDKRLMRVQGWMA 148  
Qy 129 ---LMGSRISNNPNIKETLEP-IQDLPNDPTVAKSPAKTSTALNAVNLGSDTKQSE 184  
Db 149 QDDVAGERLINS-----SAQPELAADANEADVRS-----TQDLVIPIGKIDAQQT- 195  
Qy 185 ANPYFALLESKNGKNGTPEISTSNYSYAQPVRVYDQGNSHDITVYFDGAPSGTSGKTFEY 244

Db 196 -----STVHYVTWTTDFQVDSFGQQHTLQINFRVPQT----- 228  
Qy 245 LVAMKPSDGSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKD-----LNAM 298  
Db 229 -----NNQWQATVTTDAGNVTPAVDPGTEVDQ-----TRVGVGTSQDGTGQVLLLEAA 277  
Qy 299 QAPLVN---GLPQPSANFVAGI---QPLTLDLFGKSKQONWAGAPASAAAIG----- 346  
Db 278 ANTFIVNFDNFHGLASSYDVVGANPDAGQVTRHAFTLNDQSGVITGVYSGVSLGEIGT 337  
Qy 347 -----TDIGKLPSSMPEIQTSNGNSTARNSSSTRYSQDGYPQGDLDVDTITISEG 396  
Db 338 ARNITQPAEQDIIQOL-ALAGFANQGGLEKA--GESITIKAYQDGYAMYLENFKIT--- 391  
Qy 397 KLOGKYSNSQVVDYFNIPLARFTSEDLRREGNNHYSATLDSGGPEFGPGLPOTSNGKLSV 456  
Db 392 -----YVQSNNSGIANISTSGVMGKGKLI 416  
Qy 457 NQLETSNVDMSREVMNMIIIOGFQMSKSVTTADTLMQKALELKR 502  
Db 417 GTLENSVNDLTDQFTDMIITQGFQAGAKTIQTSMTLTVLNLKR 462

RESULT 8

US-10-009-823A-10  
; Sequence 10, Application US/10009823A  
; Publication No. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Akenbauer, Robert G.  
; TITLE OF INVENTION: LAMSONIA DERIVED GENE AND RELATED FLGE  
; FILE REFERENCE: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(400)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-009-823A-10

Query Match 13.1%; Score 337.5; DB 14; Length 400;  
Best Local Similarity 25.1%; Pred. No. 6.4e-19;  
Matches 128; Conservative 69; Mismatches 165; Indels 147; Gaps 21;  
Qy 11 GKKTHSTGLGTVSNNIANANTIGYKQOVVQDLFSDOLAIGSTGSGPNQAGMAQVGS 70  
Db 21 GKVAGTLDVIGNNIANSATYGFASGTI-----TQDFTDGTITNG--TGLDVALA- 70  
Qy 71 VRTIFTOGAFEPGNSVTDLAIGKGFQO-VTLEDKVHYTRAGNFRFTQD-----GFLN 122  
Db 71 -----SFDNMFASQNGFFRLVDNSGVSFYSRNGQFKLDENRLNVNMQGIN 116  
Qy 123 ----DP-----SGFTLMGSRISNNPNIKETLEPIQDLPNDPTVAKSPAKTSTALNAVNL 174  
Db 117 LNSDPLQLQLTGYPATGT-----PPTIQOGA-----NFTVTFPSASADSNKPTNI 163  
Qy 175 GSDTDKTOSEANPYFALLESKNGKNGTPEISTSNYSYAQPVRVYDQGNSHDITVYFDGAP 234

Db 164 --SPNTLMKAK-----TITTSKMGKSVTVFDSQGNADKXVYF----- 200  
QY 235 SSTGSKTFE-YLVAMNPSGDSAGTSDAGLMSGMTWTFSSNGELKMTAFTPTGSATK 293  
Db 201 VKTGTNOKVYFGAMANNQDSSDPNSI--AKIATATTLFANAGTL-----VDTGA--- 249  
QY 294 DLNAWQAPLVNGLPOFSANFVGAGIQPLTDLDFGKSOQNNWAGAPASAAAIGTDIGKLP 353  
Db 250 -INGAEPATP--SLSNDGTVVGNYSNEQT--FLNSMQON-----TCQLL 289  
QY 354 SMPIQTSSGN--STARNGSSSTRYSQDGYPGDLVDVTTITSEGLKQKYSNQWDFYN 412  
Db 290 GQIVLANFANNEGLASEGDNIVATTQNGYKPGDLVSVQIV----- 330  
QY 413 IPLARFTSEDLRREGNNHYSATLDSGPEGLGCTSNYKGLSVNQLETSNVDMRSRMVN 472  
Db 331 -----WSATOSSGVALLTGATGNGFTLNGALEASNVDSLSELVN 371  
QY 473 MIIQRFQMNKSVTTADTLMQKALELK 501  
Db 372 MIVAQRNYQSAQRIKTQDQILNTLVNLR 400

## RESULT 9

US-10-009-823A-9  
; Sequence 9, Application US/10009823A  
; Publication No. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Ankenbauer, Robert G.  
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE  
; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES  
; FILE REFERENCE: DAV150.001APC  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Salmonella typhimurium  
US-10-009-823A-9

Query Match 12.9%; Score 333; DB 14; Length 401;  
Best Local Similarity 23.1%; Pred. No. 1.5e-18;  
Matches 124; Conservative 64; Mismatches 147; Indels 202; Gaps 17;  
QY 11 GMKTHSTGLGTVSNKIANANTIGYKQQVVFQDLFSODLAIGTSGOPNQAGQAQVGS 70  
Db 21 GKVAGTNLDVIGNNANSATYGFSGTI-----TQFTDGTITNG-----TGLDVAIA- 70  
QY 71 VRTFTQGAFFGNSVTDLAIGGKGFQ-VLEDKVHYTRAGNFRFTQDGFNDPSGFTL 129  
Db 71 -----SFADMPASQNGFFRLVDNSGVSFYSRNGQFKLDENRNLVNMQG--- 113  
QY 130 MGSRISSNPNTKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTO---SEAN 196  
Db 114 -----QINLNTDP--VMQLTGTPA\*GTPTTQOGANPSKTFPSVSDAD 155  
QY 187 PYFALLESKNGKGTPTPI-----STGNYSYAQPMRVYDQGNSHDITVYFDGAPSS 236  
Db 156 SY-----NKPATIPNTLMAAKSTTTASKMGTVTVYDVSQGNADNMVYF-----VK 202  
QY 237 TGSKTFFVLVAMNPSGDSAGTSDAGLMSGMTWTFSSNGELKMTAFTPTGSATKDLN 296

Db 203 TKDNEWAVYTGTVNHDSSDPAATAPITTTASTTLKFNENGILES-----TGT----- 250  
QY 297 ANQAPLVNGLPQFSANFVGAGIQPLTDLDFGKSOQNNWAGAPASAAAIGTDIGKLPMM 356  
Db 251 -----ING----- 253  
QY 357 PIQTSSGNSTANGSSSTRYSQDGYPGDLVDVTTITSEGLKQKYSNSQVDFYN----- 412  
Db 254 -----ATATFSL-----NNDGTVVGNYSNEQ--EFLNSMQQ 284  
QY 413 -----IPLARFTSEDLRREGNNH-----YSATLDSGGEFFG 444  
Db 285 NTGQVLGQIVLANFANNEGLASQGDNNIVATNQYKPGDLVSVQIVWAAATCASGVALLG 344  
QY 445 LPGTSNYKGLSVNQLETSNVDMRSRMVNMIIQRFQMNKSVTTADTLMQKALELK 501  
Db 345 TAGSGNFKLTNGALEASNVDSLSELVNIVAQRNYQSAQRIKTQDQILNTLVNLR 401

## RESULT 10

US-10-335-977-4818  
; Sequence 4818, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 4818:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...221  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4818:  
US-10-335-977-4818

Query Match 11.4%; Score 296; DB 15; Length 221;  
Best Local Similarity 33.2%; Pred. No. 6.6e-16;  
Matches 78; Conservative 37; Mismatches 98; Indels 22; Gaps 3;



LENGTH: 271 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...271  
SEQUENCE DESCRIPTION: SEQ ID NO: 4764;  
US-10-335-977-4764

Query Match 11.1%; Score 288; DB 15; Length 271;  
Best Local Similarity 21.0%; Pred. No. 4e-15;  
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;  
QY 1 MMGSLFIGATGKMTHTGLTVSNANIANITGYKQQQVFPQDLFSQDLAIGSTGSGPN 60  
DB 10 MLRSIYATSGMLAQOQTHIDTSSNNIANVNTGFKKSRADFNDFYQAMQVAGINTSNTT 69  
QY 61 -----QAGMGAQVGSVRTIFTQGAPEFGNSVTDLAIGCKGFFQVLTLED-KVHYTRAGNF 113  
DB 70 LSPDGMVGLGVVRPSAITKMFSGQSPKETTENNLDIAITGKGFQVQLPDGTTAYTRSGNF 129  
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKTEPIQLDFNDPTVAKSPAKTSTALNAVNV 173  
DB 130 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 159  
QY 174 LGDSTDKTQSEANPYFALLEWSKNGKNGTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA 233  
DB 160 IG-----VDGT 165  
QY 234 PSSTGSKTFEYLVANPSEDGSAAGTDSAGLLMSGTWTFSSNGELKNMTAFTPTGSATK 293  
DB 166 VSVT----- 169  
QY 294 DLNMQPAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLP 353  
DB 170 ----- 169  
QY 354 SMMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVTITSEGKLGKYSNQWDFYNI 413  
DB 170 -----QGLQTTSNVIG--QI 182  
QY 414 PLARFETSDGLREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDYSRMVNM 473  
DB 183 TLANFVNPAGLHSMGDNLFSTINASGDAIVGNPDSQGLGKLRQGLFELSNVRLVEEMTDL 242  
QY 474 IIRQGFQWNSKSVTTADTLMQKALELKR 502  
DB 243 ITAQRAYEANSKSIQTADAMLQTVNSLKR 271

## RESULT 13

US-10-335-977-4765  
; Sequence 4765, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 4765:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 279 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...279  
SEQUENCE DESCRIPTION: SEQ ID NO: 4765;  
US-10-335-977-4765  
Query Match 11.1%; Score 288; DB 15; Length 279;  
Best Local Similarity 21.0%; Pred. No. 4.2e-15;  
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;  
QY 1 MMGSLFIGATGKMTHTGLTVSNANIANITGYKQQQVFPQDLFSQDLAIGSTGSGPN 60  
DB 18 MLRSIYATSGMLAQOQTHIDTSSNNIANVNTGFKKSRADFNDFYQAMQVAGINTSNTT 77  
QY 61 -----QAGMGAQVGSVRTIFTQGAPEFGNSVTDLAIGCKGFFQVLTLED-KVHYTRAGNF 113  
DB 78 LSPDGMVGLGVVRPSAITKMFSGQSPKETTENNLDIAITGKGFQVQLPDGTTAYTRSGNF 137  
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKTEPIQLDFNDPTVAKSPAKTSTALNAVNV 173  
DB 138 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 167  
QY 174 LGDSTDKTQSEANPYFALLEWSKNGKNGTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA 233  
DB 168 IG-----VDGT 173  
QY 234 PSSTGSKTFEYLVANPSEDGSAAGTDSAGLLMSGTWTFSSNGELKNMTAFTPTGSATK 293  
DB 174 VSVT----- 177  
QY 294 DLNMQPAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLP 353  
DB 178 ----- 177  
QY 354 SMMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVTITSEGKLGKYSNQWDFYNI 413  
DB 178 -----QGLQTTSNVIG--QI 190  
QY 414 PLARFETSDGLREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETNSVDYSRMVNM 473  
DB 191 TLANFVNPAGLHSMGDNLFSTINASGDAIVGNPDSQGLGKLRQGLFELSNVRLVEEMTDL 250  
QY 474 IIRQGFQWNSKSVTTADTLMQKALELKR 502  
DB 251 ITAQRAYEANSKSIQTADAMLQTVNSLKR 279  
RESULT 14

US-10-009-823A-12  
; Sequence 12, Application US/10009823A  
; Publication No. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Ankenbauer, Robert G.  
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE  
; FILE REFERENCE: DAVI150.001APC  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 360  
; TYPE: PRT  
; ORGANISM: Vibrio parahaemoliticus  
US-10-009-823A-12

Query Match 9.0%; Score 233; DB 14; Length 360;  
Best Local Similarity 22.0%; Pred. No. 1.9e-10;  
Matches 110; Conservative 59; Mismatches 127; Indels 204; Gaps 21;

QY 11 GVKTHSTGLTVSNINANANTIGYKQOVVFDLFSODLAIGSTGSGPNQAGMAQVGS 70  
DB 20 GVEVASTELNISHIANASTYGFK-----GARI5Q 50

QY 71 VRTIFQGAFFPGNSVT-----DLAI-----GGKGFQVTLK-----DKVHYTRAGNFR 114  
DB 51 -----NFKNGSITGTGRMDLAITEFAAVYNNNGSGFF-VTKDHMGOTLYTRSGVFG 101

QY 115 FTQDGLNDPSGFT-----LMGSRISNNPNKIKETLEPIQLDEN--DPTVAKSPA 162  
DB 102 TDKSNFVTANNNGVANFNDASAKAKLQGYSDSNNNLMTGSDKAVTPDPADP-----154

QY 163 KTSTALNAVVALGSDTDKTOSEANPYFALLESWKNGTPPITSTSNYSYAQPMRVYDOGN 222  
DB 155 --TSFNSVGNIQVSTSLNKA-----TDKLDPSYTTQVYDSLGN 192

QY 223 SHDIIVP-----DGA-----PSST-----GSKTPEYLVANMP-----SEDSASGTD5AGL 265  
DB 193 SHTVQYFTKTADNAWEVNVPTGSEFNQVDGKGT--PVVSTIPVTNKDGTLLAAFPAPAG- 249

QY 266 LMSGTMTFSSNGELKNMTAFTPTGSATKDLNAPLVLNGLPQFSANFVGAGIQPLTLD 325  
DB 250 --ANAMSVVDINEDNGMVVATYTNQSLK-----GSTQFGAQLQG--QVVLAD 292

QY 326 FGKSKQNMWAGAPASAAAIGTDIGKLP5MMPITQSSNGSNARNGSSSTRYSQDGYPOG 385  
DB 293 FA-----NTQGLAKVSGFGVSTN-----SPNGYTS 318

QY 386 DIAVDYITITSEGLKQCKYNSQWVDVFNIPLARFTSDDLREGNHNSATLDSGGPEFGL 445  
DB 319 ELAGVRVA-----WTQSSSGAPIMGV 340

QY 446 PGTSNYGKLSVNQLETSNVD 465  
DB 341 PGSGTGLNLTGALGSGNVD 360

RESULT 16  
US-10-335-977-4840  
; Sequence 4840, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:

US-10-009-823A-11  
; Sequence 11, Application US/10009823A  
; Publication No. US20030157120A1  
; GENERAL INFORMATION:  
; APPLICANT: Panaccio, Michael  
; APPLICANT: Rosey, Everett Lee  
; APPLICANT: Sinistaj, Meri  
; APPLICANT: Hasse, Detlef  
; APPLICANT: Parsons, Jim  
; APPLICANT: Ankenbauer, Robert G.  
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE  
; FILE REFERENCE: DAVI150.001APC  
; CURRENT APPLICATION NUMBER: US/10/009,823A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: PCT/AU00/00437  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/133,973  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 480  
; TYPE: PRT  
; ORGANISM: Aquifex aeolicus  
US-10-009-823A-11

Query Match 10.9%; Score 282; DB 14; Length 480;  
Best Local Similarity 22.9%; Pred. No. 2.9e-14;  
Matches 135; Conservative 73; Mismatches 160; Indels 222; Gaps 28;

QY 1 MMSLFTGATGM-----KTHSTGLTV-----SNINANTIGYKQOVVFDL 44  
DB 1 MLRSFYNAITGMDVSRGTVKTITFGAGAVVDSFALDVTSDNLNANVTGVFKSR-----54

QY 45 FSODLAIGSTGSGPNQAGMAQVGSVRTITGAFEPGNSVTDLAI-----91  
DB 55 -----TQKVITGSGFKQTEBITTDLAIPFDQMV5QWVVG 88

QY 92 -----GKGF--QVTLKDVHTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNKIKET 144  
DB 89 LNTTTEGALFILRDVLNQTYYTRDRGRFRINREGYLINFG-----130

QY 145 LEPIQLDFNDPTVAKSPAKTALNAVNLGSDTDKTOSEANPYFALLESWKGN-----198  
DB 131 --PFLDERAPILYVQGFKNVPTGEVT--GTQIDQTTTFNP-----LDSFTNYLEDIR 182

QY 199 ---GTPPISTSNYSYAP---MRVYDQGNSHDITVYFDGAPSGT-----238  
DB 183 VETQIPPKATGEI--YFNPRYTLTIYDSLGREVPADIYF---VKITGNQWKVYFTLPTFA 237

QY 239 SKTPEYLVANMP5EDGSA-----SGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGS 290  
DB 238 SKTLF-----DPSLASKERYINVDWNGDDDKGLV-----YIPGD 276

QY 291 ATKDLNAWQAPLVNGLPQFSAN--FVAGIQPLTLDFGIKSQNMWAGAPASAAAIGT- 347  
DB 277 IVQD-----TANQKFLYTDIVELDL--FNDQVHI-----ADNGTF 309

QY 348 -DIGKLP5MMPITQSSNST-----ARNGSSRFR-RYSQDG-----YPOGDLVDV-- 390  
DB 310 SEVDLTPESGFSEINDPNDTESYLNKLGAKLGETNKKLSEGVVGVVYSNGETIYVGE 369

QY 391 TITSEGLKQKYSNSQWVDVFNIPLARFTSDDLREGNHNSATLDSGG-----440  
DB 370 GILQNNVQNSYITQHALPYRLAQAQTPDEELVKKGS-D-FVVTNDQGVARGELIDL 428

QY 441 -----PEPGLGTSNYGKLSVNQLETSNVDMSREMNMIIOQGFQM 482  
DB 429 VLYASVKTPTILLPGGSGN--KIRSAVVMGNSVDIAKEFINLITACQTYOV 476

RESULT 15



FILE REFERENCE: DAVI150.001APC  
CURRENT APPLICATION NUMBER: US/10/009,823A  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: PCT/AU00/00437  
PRIOR FILING DATE: 2000-05-11  
PRIOR APPLICATION NUMBER: US 60/133,973  
PRIOR FILING DATE: 1999-05-13  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 370  
TYPE: PRP  
ORGANISM: Borrelia burgdorferi  
US-10-009-823A-13

Query Match 7.2%; Score 186; DB 14; Length 370;  
Best Local Similarity 21.9%; Pred. No. 1.4e-06;  
Matches 98; Conservative 57; Mismatches 143; Indels 150; Gaps 19;

QY 1 MMSLFTGATGKTH-----STGLGTVSNINANITGYKQOQVWFQDLF 45  
DB 1 MMSLYSGVGLQHOGTNPQVGLGNVASTMDVVGNNIANVNTIGFKK-----51  
QY 46 SQDLAIGSTGSGQPNQAGMAQVGSVRTITFQAGFEPGNSVTDLAI-----91  
DB 52 -----GRDITHTQAGAFQSTQKASDLGVVNFQDMISQSIGA 88  
QY 92 -----GKGFQFVLEDKVHYTRAGNFRFTDQGLNDP-SGFTLMGSRISNNPNIKK 142  
DB 89 SRPTDARGSGNGFFILKEGFLNFTTRAGAFDVSDRHLVNPANGACNLDKRLVRIQGWMA 148  
QY 143 ETLPEIQL--DFNDPTVAK--SPAKTSTAL-NAVNLGDSSTKTOSEANPYFALLBSWK 197  
DB 149 RDLGKRVINTASDPLIOEGANPADIARGIEDLIIPIDGREGAKSTKGVTF-----TWVV 203  
QY 198 NGTPISTSYSAQPMRVYD-QQGNSHDITVYFDGAPSTGSKTFEYLVAMNPSEDGSA 256  
DB 204 NKSLYDSFGVSVLELRVWLDLNTPLNWNATV-LNGQP-----LNGEONG-- 251  
QY 257 ASGTDSAGLMSG-TMTFSSNGELKMNFTPTGSAATKOLNAPQAPLVNGLPQFSANFV 315  
DB 252 ----DILQIPISNFTLGFNEGALAS-----TFNVL 278  
QY 316 GAGIQPL--TLDFGKISQNNWAGAPASAAAIGT-----DIGK--LFSMMP 357  
DB 279 GANVTEGETQTVNLKDDQGVIVIGYNGIRLGTGVSYTDSITQFADRDILGKIALAFM- 337  
QY 358 IQTSSGNSTARNSSSTRYSQDGPQ 385  
DB 338 ----NPGDLQNALSSTKAILQDGYGMG 361

RESULT 19  
US-10-369-493-4713  
Sequence 4713, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 4713  
LENGTH: 655  
TYPE: PRP

ORGANISM: Burkholderia fungorum  
US-10-369-493-4713  
Query Match 7.0%; Score 180.5; DB 14; Length 655;  
Best Local Similarity 22.4%; Pred. No. 8.9e-06;  
Matches 138; Conservative 75; Mismatches 225; Indels 178; Gaps 30;

QY 1 MMGSLFIG-ATGMKTHSTGLGTVSN--IANANTIGYKQOQVVFQ-----DLFQD 48  
DB 98 LIGSPTAGIASAITSYFIFGLQNVSNASSLATRQTAMSGAQLVQVINAAGQOYDALRQS 157  
QY 49 L--AIGSTGSGQPNQAGMAQV-GSVRTIFFTQAGFEPGNSV--TDLAIGKGFQVTLT 102  
DB 158 VNTQLTNTVQSINSYSQOIAQLNGQIAQASTQG--QPPNQLLDQDLAVSN-----206  
QY 103 DKVHYTRAGNFRFTDQGLNDPSPFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKPSA 162  
DB 207 -----LSQLIGVNVVNS--NGSYVFMGSGPLVSTSYNIGTAPS 246  
QY 163 KTSTALNAVNLGDSSTKTOSEANP-----YFALLBSWKNGTPTPI 203  
DB 247 TGDTSLSVOYLG-----QAGANPAATPQNLPSDKIDGTLGLGLLAFRSQTLDPGEAQL 300  
QY 204 STSNYSYAQPMRVYDQGNSHDITVYFDGAPSTGSKTFEYLVAMNPSEDGSAASGTDSA 263  
DB 301 GAIAVSFS--AQVNAQNGLGLITLAGAKGALFSVGGPT-----VYANTQNTGNASUNVSFA 354  
QY 264 GLLM--SGTMTFSSNGELKMNFT--AFTPTGSAATKOLNAPQAPLVNGLPQFSANFVGAG 318  
DB 355 DATQPTTGDTYLAINGTITTYLTLDNSTGTIVGSAT--NLSQP-----INGL--NFSITG 403  
QY 319 IQPLDLDFGKISQNNWAGAPASAAAIGTIDGKLPSPMMP-----QTSNGSTARNSS-- 371  
DB 404 ----TNWAGDSFTVETFRGALNSFATATTDASIAAAPVLAASNTGTATITQGTVT 459  
QY 372 -----SSTRYSQD-----GYEQGLVDV-----TITSE 395  
DB 460 AGYTPNPTTLLSYDGTGLSGFPAGSTTVVAGTPPTTITASAATVVPYSATGATLIN 519  
QY 396 GKLOGKYSNVQVDFYNIPLA--RFT--SEDLRREGNN-----HYSATLDSGGPEFGLP 446  
DB 520 NATAGQNNVSVT--ISNAPAGDKFTIGTNTGATNDGRNALALSNTAKAMSGGTVTILT 578  
QY 447 G-----TSNYGKLSVNOLETNS-----VDMSERVMNMIIOGFGM 482  
DB 579 GAYANYVNNIGN-QTNQIQTSSAOSLSVQTITTAQOSVSGVNVINEEAANLLQYQOLYQA 637  
QY 483 NSKSVTTADTLMQKAL 498  
DB 638 NSKVIQTAQTLFQTIL 653

RESULT 20  
US-10-369-493-7473  
Sequence 7473, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 7473  
LENGTH: 655  
TYPE: PRP



Db 1011 IKAGVDGENSDSDATNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059  
Qy 71 VRTFTQGAPEPGNS-----VTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLND 123  
Db 1060 DLTIGNTNSADGNVAKVFNQVDKSIADG-HKVTLHSKVE--TSGSNNTEDSSDNN 1116  
Qy 124 PSQFTLMGSRISNNPIKKELEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTQS 183  
Db 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGBITTKTGTINATTGNVEITATQGS 1170  
Qy 184 EBNPYFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238  
Db 1171 -----ILGGIESSSGSVTLTATEGALAVSNIS-----GNTVVTIAN--SCALTTLA 1214  
Qy 239 SKTFEYLVANMPSEDSGAASCTDSAGLL--MSGTMTFSSNGELKNMTAFTPTGSAKDL 295  
Db 1215 GSTIKGFEVTSQSQSDIGGTISGGIVKATESLTQSNISKATTEANVTSAITGI 1274  
Qy 296 NAMQAPLVNGLPQFSANFVAGIQLTLDLFGIKSQNMWAGAPASAAIGTDIGKLPSP 355  
Db 1275 GGTISGNTVN-----VTANAGDLTVGNAGAE-----INATEGAATLTSSGKL---- 1316  
Qy 356 MPIQTSSGNSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL-----OGK----- 401  
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAAVNTLTGTLTVKGSNNINATSGTL 1372  
Qy 402 YSNSQVDFNIPILARTSEDLRRGNHYSATLDSGGPEFGPGTSGNYKLSVNQLET 461  
Db 1373 VINAKDAELNGAALGNHTVVNATNANGSGSVIATTS--RVNITG----DLITINGL-- 1423  
Qy 462 SNVDMREVMNMIIOGFQFQNSKSVTTADTLMQKALELKR 502  
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEIVEAKR 1461

RESULT 24  
US-10-193-764-63  
; Sequence 63, Application US/10193764  
; Publication No. US2003013943A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 77944  
; LENGTH: 3705

Query Match 6.7%; Score 173.5; DB 14; Length 1536;  
Best Local Similarity 22.8%; Pred. No. 0.00012;  
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;  
Qy 12 MKTSTGLTGVSNINIAN-TIGYKQOVFQDLFSQDLAIGTSGSQPNQAGVQGS 70  
Db 1011 IKAGVDGENSDSDATNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059  
Qy 71 VRTFTQGAPEPGNS-----VTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLND 123  
Db 1060 DLTIGNTNSADGNVAKVFNQVDKSIADG-HKVTLHSKVE--TSGSNNTEDSSDNN 1116  
Qy 124 PSQFTLMGSRISNNPIKKELEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTQS 183

Db 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGBITTKTGTINATTGNVEITATQGS 1170  
Qy 184 EBNPYFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238  
Db 1171 -----ILGGIESSSGSVTLTATEGALAVSNIS-----GNTVVTIAN--SCALTTLA 1214  
Qy 239 SKTFEYLVANMPSEDSGAASCTDSAGLL--MSGTMTFSSNGELKNMTAFTPTGSAKDL 295  
Db 1215 GSTIKGFEVTSQSQSDIGGTISGGIVKATESLTQSNISKATTEANVTSAITGI 1274  
Qy 296 NAMQAPLVNGLPQFSANFVAGIQLTLDLFGIKSQNMWAGAPASAAIGTDIGKLPSP 355  
Db 1275 GGTISGNTVN-----VTANAGDLTVGNAGAE-----INATEGAATLTSSGKL---- 1316  
Qy 356 MPIQTSSGNSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL-----OGK----- 401  
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAAVNTLTGTLTVKGSNNINATSGTL 1372  
Qy 402 YSNSQVDFNIPILARTSEDLRRGNHYSATLDSGGPEFGPGTSGNYKLSVNQLET 461  
Db 1373 VINAKDAELNGAALGNHTVVNATNANGSGSVIATTS--RVNITG----DLITINGL-- 1423  
Qy 462 SNVDMREVMNMIIOGFQFQNSKSVTTADTLMQKALELKR 502  
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEIVEAKR 1461

RESULT 24  
US-10-282-122A-77944  
; Sequence 77944, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 77944  
; LENGTH: 3705

```
; TYPE: PRT
; ORGANISM: Versinia peetis
US-10-282-122A-77944

Query Match
Best Local Similarity 6.7%; Score 172.5; DB 15; Length 3705;
Matches 119; Conservative 58; Mismatches 180; Indels 191; Gaps 24;

QY 3 GSLFTGATGKTHSTGLCTVSNNTANITGYKQOVVFDLFSQDLAIGTSGSQGNQA 62
Db 2359 GELIIGAGAGCQTSLLNIAG--ASANINGRQ-----TVGAVTNGVTLL 2403
QY 63 GNGAAGVSVRIFTQGAPEPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLN 122
Db 2404 GNGBELTSTDILNTGMINVTGDLNLENGGASSISGGL-----TGNGILN 2449
QY 123 DPSG-FTLMGSRISNNPNIKETLEPIQLDFNDFTVAKSPAKTALNAVNLGSDTKT 181
Db 2450 IKGSDFTI---SIDNN-----GLAGQTNISDGASVT 2477
QY 182 QSEANPYFALLESKWNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTSGKT 241
Db 2478 -----LNGGTIIIGCN-----LGSSV 2494
QY 242 FEYLVAMPSEDGSAAS-----GT--DSAGLLMSGTWTFS-----SNGEL-----KNMT 283
Db 2495 IDVLGDLNLVADNSLANVISGDGTINTTATVTLSGNSFSFGAHQIGTNGELTVGQASNLG 2554
QY 284 AFTPTGSAKDNLNAQWAPLNGLPQFSANFVGAGIQPLTLD----- 325
Db 2555 A-----SSATVNLGLTSLHILNGVSESIAN-VLSGVAGSVTDIIGGADTALTANNSGFLG 2609
QY 326 -----FGIKSQONWAGAPASAAAIGTDIGKLPMSMPIQTSSGNSTARNSSSTRR 376
Db 2610 QVALAGNSKLTAVASTNUL--GASSVALAGT--GDTLSLGFNGTFCNSVTGSGVLQVT- 2664
QY 377 YSQDGYPGQDLVDVTIITSEGLQKGYNSQVDFNIPL-----ART---SEDGLREG 428
Db 2665 -----DRAEVLTSNGV---SNAVTDIADATLNLDDIALFNHALTGNNLVA 2711
QY 429 NNHYSATLDSGPFEG-LPGTSNYKLSVNOLETNSVDMGSRVMNIIIQGFQMSKSV 487
Db 2712 KNDASTAFDFGATVGGAFTGVN-----LNNSTFDLSGNNTT-VLAQATLKLSGNL 2762
QY 488 TTADTMLQ 495
Db 2763 TSVGNQVQ 2770

RESULT 25
US-10-335-977-4817
; Sequence 4817, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977

; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4817:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...91
; SEQUENCE DESCRIPTION: SEQ ID NO: 4817:
US-10-335-977-4817

Query Match
Best Local Similarity 6.4%; Score 166; DB 15; Length 91;
Matches 33; Conservative 20; Mismatches 33; Indels 0; Gaps 0;

QY 400 GKYSNSQVDFNIPLEAFTSEDEGLREGNNHYSATLDSGPFGLPQTSNYKLSVNQL 459
Db 5 GAFSGRTIALAQAVALANFANDAGLQALGDVNFQGTGNSGOALIGAANTGRRSIGSKL 64
QY 460 ETSNVDMSGRVMNIIIQGFQMSK 485
Db 65 ESSNVDLSRLTNLIVQGFQKTLK 90

RESULT 26
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match
Best Local Similarity 6.4%; Score 166; DB 14; Length 2468;
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

QY 20 GTVSNNTANITGYKQOVVFDLFSQDLAIGTSGSQG-----PNQAGMAQVGSV 71
Db 1298 GTVNVAVA-----QDPA-GNTGPGSTTVDVAFNPTVNVNPSNGL 1337
QY 72 RTITQGAPEPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLNDPSGFTLMG 131
Db 1338 ----LNGTAEFGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSQLPNGTVVAVTA 1387
```

```
QY 132 SRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFAL 191
Db 1388 SDAAGNTSLPATT---TVDSSLPSIPQVDPSPNGSVISGTADAGNTIIITDNGNPIQGV 1443
QY 192 LESWKNG--TPPISTNSYSAQPMRYVDOQGNSHDITVYFDG-APSSGSKTFFYLAVM 248
Db 1444 TADGSGNWSTFGIPLPDGTGVNVVARSNVDSPAPVITVDGVAPAP---VI 1494
QY 249 NPSEDGSAASGTDAGLLM-----SGTMFTS-----SNGELKNMTAFT 286
Db 1495 DPS-NGTEISGTAEAGATVILTDGGNPIGQATADGSGNWTTPSTPLANGTVINAVAQD 1553
QY 287 PTG-----SATKLNWQAPLNV-----GLPQFSANFV-----GAGIQPLTLDG 327
Db 1554 PAGNTSGPASVTVDAIA-PPAPVNPNSGVVISGTAEAGATVILTDGNGNPIQGV 1610
QY 328 IKSQONWAGAPASAAAGTDIGKL-----SGTMFTS-----SNGELKNMTAFT 367
Db 1611 ---GSGNWSTFGTPLANGSVINALQAAGNNSPTSATVDSLAPAPVIDPNSGVIA 1667
QY 368 RNGSSSTRYQD--GYQGDLDVDTITSEKLGKQKYSQVDFYNIPLARFTSEDGLR 425
Db 1668 GTAEAGATVILTDGNGNPIQ-----VTADG--SGNWSFTPGTPLSLNGTVVNAVAQDA-- 1718
QY 426 REGNNH--YSATLDSGGPEGLPGTSNYGKLS 455
Db 1719 -AGNTSGPVSTTVDAVAPATFVIDPNSGVELS 1749

RESULT 27
US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match 6.4%; Score 166; DB 15; Length 2468;
Best Local Similarity 21.9%; Pred. No. 0.00095;
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

QY 20 GTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGQ-----PNOAGMCAQVGSV 71
Db 1298 GTVNAVA-----QDPA-GNTPGQSTTVDAVAPNTPVVPNPSGNL 1337
QY 72 RTITQGAPEPNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQGLNDPDSGFTLMG 131
Db 1338 ---LNGTAEPGSEVTITLTDGNGNPIGQTADG-----SGNWSFTPGSOLPNGTVVNTA 1387
QY 132 SRISNNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFAL 191
Db 1388 SDAAGNTSLPATT---TVDSSLPSIPQVDPSPNGSVISGTADAGNTIIITDNGNPIQGV 1443
QY 192 LESWKNG--TPPISTNSYSAQPMRYVDOQGNSHDITVYFDG-APSSGSKTFFYLAVM 248
Db 1444 TADGSGNWSTFGIPLPDGTGVNVVARSNVDSPAPVITVDGVAPAP-----VI 1494
QY 249 NPSEDGSAASGTDAGLLM-----SGTMFTS-----SNGELKNMTAFT 286
Db 1495 DPS-NGTEISGTAEAGATVILTDGGNPIGQATADGSGNWTTPSTPLANGTVINAVAQD 1553
QY 287 PTG-----SATKLNWQAPLNV-----GLPQFSANFV-----GAGIQPLTLDG 327
Db 1554 PAGNTSGPASVTVDAIA-PPAPVNPNSGVVISGTAEAGATVILTDGNGNPIQGV 1610
QY 328 IKSQONWAGAPASAAAGTDIGKL-----SGTMFTS-----SNGELKNMTAFT 367
Db 1611 ---GSGNWSTFGTPLANGSVINALQAAGNNSPTSATVDSLAPAPVIDPNSGVIA 1667
QY 368 RNGSSSTRYQD--GYQGDLDVDTITSEKLGKQKYSQVDFYNIPLARFTSEDGLR 425
Db 1668 GTAEAGATVILTDGNGNPIQ-----VTADG--SGNWSFTPGTPLSLNGTVVNAVAQDA-- 1718
QY 426 REGNNH--YSATLDSGGPEGLPGTSNYGKLS 455
Db 1719 -AGNTSGPVSTTVDAVAPATFVIDPNSGVELS 1749

RESULT 28
US-10-335-977-4839
; Sequence 4839, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4839:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...124
; SEQUENCE DESCRIPTION: SEQ ID NO: 4839:
US-10-335-977-4839

Query Match 6.4%; Score 165.5; DB 15; Length 124;
Best Local Similarity 32.0%; Pred. No. 1.3e-05;
Matches 39; Conservative 24; Mismatches 48; Indels 11; Gaps 3;

QY 392 ITSEGLQKQYNSQVDFYNIPLARFTSEGLRREGNNHY---SATLD-----SGGPE 442
Db 3 IEENGVISLAFSNGVVEPARIGIXAFINDGRLKIGLNLYEMEGFTINGENPLXGNPI 62
QY 443 FGL--PGTSNYGKLSVNOLETSVNDMSREVMNIIIOGFQMSKSVTTTADTLMOKALEL 500
Db 63 LGWDEEGLKFGKIRHKYLETSNVNAGNALTLLIMQGYSMNARAFGAGDDMIKEAISL 122
QY 501 KR 502
Db 123 KK 124

RESULT 29
US-10-335-977-4819
; Sequence 4819, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; City: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018

; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4819:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...343
; SEQUENCE DESCRIPTION: SEQ ID NO: 4819:
US-10-335-977-4819

Query Match 6.3%; Score 164; DB 15; Length 343;
Best Local Similarity 24.5%; Pred. No. 7.7e-05;
Matches 82; Conservative 52; Mismatches 129; Indels 72; Gaps 15;

QY 32 IGYKQQVVFQDLFSODLAIGS-----TGSQGN--QAGHGAQVGSVRTIPTGGAPEPGNS 85
Db 3 LGFKYSRASFDMLSQVKLIATDPYKNGLAGQNDPSVGLGVGVDTATTKIFSCGNIQNTDV 62
QY 86 VTDLAIGKGFQVTLDEKV--HYTRAGNFRF---TQGFNLDPGFTLM-----GSRI 134
Db 63 KTDLAIQDGFIIISPDGRTNFTRAGVSGFLAREFGYHRWACGARSVENRSDTGNGK 122
QY 135 SNNPNIKETLEPIQLDNDPTVAKSPAKTST--ALNAVNLGSDTKTQSEANPYFALL 192
Db 123 SPTDALKVDNTGPLENIRIDPGMV-MPAKTSNRISMRIANINAGKHADQTA--VFALD 177
QY 193 ESWK--GNGTPTISNTSYAQPVR---VYDQGN-----SHDITVYFDGA----- 233
Db 178 SSAKTPSDGINPVYDSGTNLAHVPENMGSLYNEGDALLLIENHGIWVYSKAKVXDIL 237
QY 234 PS-----STGSKTFEYLVMNPSEDSGAAG-----TDSAGLLMSGTM 271
Db 238 PSAENSTLELNGVKISFTNDSAVSRSTSLVAAKNAINAVKSGTGTIEAYLDGKQLRLNTN 297
QY 272 TFSSNGELKNMT-----AFTPTGSATKDLNAWQ 299
Db 298 ELDGKELKNIVTQAGTGAFANFLDGDQKVTAFK 332

RESULT 30
US-10-147-299A-4
; Sequence 4, Application US/10147299A
; Publication No. US20040058323A1
; GENERAL INFORMATION:
; APPLICANT: KO, ALBERT I.
; APPLICANT: HAAKE, DAVID A.
; APPLICANT: REIS, MITERMAYER GALVAO
; APPLICANT: MATSUNAGA, JAMES
; APPLICANT: CRODA, JULIO HENRIQUE ROSA
; APPLICANT: SIQUEIRA, ISADORA CRISTINA
; APPLICANT: RILEY, LEE W.
; APPLICANT: BAROCCHI, MICHELE
; APPLICANT: YOUNG, TRACY ANN
; TITLE OF INVENTION: PROTEINS WITH REPETITIVE BACTERIAL-IG-LIKE (BIG)
; FILE REFERENCE: 3673-19
; CURRENT APPLICATION NUMBER: US/10/147,299A
; FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1954
; TYPE: PRT
; ORGANISM: Leptospira kirschneri
US-10-147-299A-4
```

```

; ORGANISM: Helicobacter pylori
US-09-881-752A-150
; Query Match 6.3%; Score 162; DB 15; Length 1954;
; Best Local Similarity 19.3%; Pred. No. 0.0014;
; Matches 125; Conservative 80; Mismatches 241; Indels 200; Gaps 23;
;
QY 7 IGATGKTHSTGLG-TVSNANANA-----NTIGYKQOQVFPQDLFSQDL 49
Db 410 VGSTNTAAIGVDITLNLVNTALLESIOVSDSHSIARGTSTFQALGVYSDGSSQNI 469
QY 50 -----AIGSTGQGPNAQMGAGVGVRTITQGAFFPGNSVTDL 89
Db 470 SDQVAMNNSNSSLQISNLNAVPKREIQSPSSGLG-----TARITATLEAISYTDI 522
QY 90 AICGKFFQVLTEDKVHYTRAG-NPFTQDGFNDPSGFTLMGSRISNNPNIKKTEL- 145
Db 523 SVNAATLVSEIPTSNPSSVSGUTVFTATGVTDGNSONLISQVTVNSNTNRATISNA 582
QY 146 ---EPIQLDFNDPT-----VAKSPAKTSTALNAVVN-----LGDSTDKT- 181
Db 583 NGTQGLAGSSVCTTNI SATLGAVTSSATLTVTNAVLNSITITPSPSVAVGSRSLNLT 642
QY 182 ---QSEANPYFALLESWKGNGTTPPISTSNYSVAQPMRVYDQGNSE----- 224
Db 643 TGYSDGNSQDLTSAVMTSDSIVSDNASGRQGTGVAQGNTOISATLGGTSSAIN 702
QY 225 -----DITYYFEGAPSTGSKTFEYLVAWNPSEDGSAASGTDSA----- 263
Db 703 FTVSAAVLDSIQVLEDSPTAKGTST--RAIATGVFSDGSLNLSIDQVWDSSTQNVQL 760
QY 264 GLLMSG-----TMTFSSNGELKXMTATPTGSGAT-----KDLNAWQPA----- 301
Db 761 GVLETPGKKLMSNPANG---NSTTGTSTRITATLGGVSGYADLTVAPSLTISQIDPHT 817
QY 302 PLVNLGPOFSANFVGAGIQPLTLDGFKSQOQNM-----WAGAPASAAAGTDIGKLP 356
Db 818 SVANGLTQ---NFATQGV-----YSDGNSQNLTDSTVWASSNPATVATSNAGTNGKAT 868
QY 357 PIOTSSGNSTARNGSSSTRYSQDGYPGDLVDVTITSEGLQCKGYSNOVDVFNIP 416
Db 869 TLQGTGNTISASLGATFS-----DPSVLTVT-----NATLTSITTAPTS 907
QY 417 RFTSEDGLRBE-----GNHYSATLDSGGGPEGLPGTSGVGL 454
Db 908 SFNIAKGLNQDFVATGYTTPGSSRDLTQTVTWNSNTSTATTISNANGTQGRMAAVDTGST 967
QY 455 SVN-----QLETNSVDMSREMNVMIIQRFQWNSKSVTTADTML 494
Db 968 NISASLGTYTQTTNVTVTSAVLNSIQVSPADISVAKGNTKAYTAI 1013

RESULT 31
US-09-881-752A-150
; Sequence 150, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Comen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 1230
; TYPE: PRN

```

```

; ORGANISM: Helicobacter pylori
US-09-881-752A-150
; Query Match 6.2%; Score 160; DB 9; Length 1230;
; Best Local Similarity 21.6%; Pred. No. 0.0011;
; Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;
;
QY 20 GTVENNIANANTIGYKQOQVV--FQDLFSQDLAIGS-TGSGPNQAGMGAGVGVRTIFT 76
Db 542 GTTINTQAKSNAPKLKAMVVVNNNEEAKTANLAQSSGTTTQSPNSTVNGA-----LNTVL- 596
QY 77 QGAPEPGNSVTDLAIGKGFQVLTEDKVHYTRAGNERFT-QDGFNDPSGFTLMGSRIS 135
Db 597 -----QNVSNFQQSIOAFOQESNIOAWANAIY 625
QY 136 NPNNIKETLEP-----IQDLFENDPTVAKS-PAKTSTALNAVYNLGDSTDKTQSBANP 187
Db 626 NTNCSQSEMTFNNNQDLRIQLRANFYQLINTINQVPTDMNALINQSQQTQOTSGSASN 685
QY 188 YFALLESKMGNGTTPPISTSNYSVAQPMRVYDQGNSHDITVYFPCAPSSTGSKTFEYLVA 247
Db 686 NNACASGMSG-----SNGNWCYQQ-----WSDSKAYYSGLQSALGYQT----- 723
QY 248 MNPSEDGSAASGTDGAGLLMSGT--MTFSSNGELK-----NMTAFTPTGSGATK 293
Db 724 ---QATTQSGSGNGNSITVNVQOITLTSNGLLNQITNLKSVNGNGASGTSGENGTS 778
QY 294 DLNAWQPAFLVNLGPOFSANFVGAGIQPLTLDGFKSQOQNMWAGAPASAAAGTDIGKLP 353
Db 779 QIN-----TAYQMLT-----DASDGKLT----- 797
QY 354 SMPQTQSSG-----NSTARNSSSTR-RYSQDGYPGDLVDVTITSEGLQCKGYSN 404
Db 798 ---YSSSSGNGNGVTPCNSST--NGSNKTSNCCVPEKNQONATATATTSNLSQKVYND 851
QY 405 SOVDVFNIPLARFTS---EDGLRR--EGNNHYSATLDS--GGPEFGLPGTSTYVGL--S 455
Db 852 AQKI--ANLIASSGNKGVENGKQFFEALKNNSSLSNLCNGSSGSGGTTCGWLNLN 909
QY 456 VNQLTSNVDSREMNVMII---IQRFQWNSKSVTTADTMLQKAL 498
Db 910 LGAIPTNGVSDTNNLLNLTETFIKTAGFIQNNDSVSTSLTSAFOAI 956

RESULT 32
US-10-282-122A-58993
; Sequence 58993, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, R.
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06

```



Db 2555 GAVGERQIINVANATNSTDAVNLSQLOAMGANVNSGVVTVNAFVAVDDSTGKVTUGGV 2614  
QY 353 PSMPFI-----QTSSGNSTARNSSSTRYSQDGYPOGDLVDV-----TITSEKLOG 400  
Db 2615 GSSMPVTLNVAEGQVTSQQAINGSQLYGTANSVASALGCTSSVSGNNGVNTAAFLDG 2674  
QY 401 KYNSN--QVVDYFNIPLARFTSEGLREGNNHYSATLDSGGPEFGLP-----GTSN 450  
Db 2675 KYNSVATMDALNAKIAT--GSTDGVVDTSAHKNLTL--GGVATTPVTVANVAATS 2731  
QY 451 YKLSVNOLETS--NVDMSREVMNMII 475  
Db 2732 DOAVNLALQKAAGLVNDSGNVTSFV 2758  
RESULT 34  
US-10-369-493-13825  
; Sequence 13825, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 13825  
; TYPE: PRT  
; ORGANISM: Pseudomonas fluorescens  
US-10-369-493-13825

Query Match 5.9%; Score 153.5; DB 14; Length 679;  
Best Local Similarity 20.3%; Pred. No. 0.0015;  
Matches 127; Conservative 69; Mismatches 208; Indels 221; Gaps 30;  
QY 5 LFIGATGKTHSTGLGTVNNIANANTIGYKQQQWVFQDLFSQDLAIGTSGSQGNQA-- 62  
Db 2 LNIGMSGLAAGSSSLAVTGNNIANVDTAGYSRQTV-----GGTKSSIQ 45  
QY 63 -----GNGAQVSVRTFTQGAPEPGNSVTDLAIGKGF--FQVTLKDYHYTRAGNFRF 115  
Db 46 YGNVFIGTGLADVRRVYNSYLESQLRTATSLNSEAAAFGAQATALD-----GSLSD 98  
QY 116 TDGFLNDFSGF--TLMSGRISNNPNIKETL-----EPIQLDFNDPTVAKSPAKTSTALN 169  
Db 99 TWTGLTGLVLEKFTSMQGVSTSDTDSRSQSVLTGAQALTSRFN--ALAKQMDQXATLN 156  
QY 170 AVNVLGSDTDKTSQSANPYFALLSWKNGTPPTISTNYSVAQPMRYD----- 218  
Db 157 G--NLSDVASQVKNLATS--IATLNQKIGB-----ISTSG--GQPNLLDSRNEAVRQLSE 206  
QY 219 -----QQGNSHDI-----TVYFDGAPS-----STGSKTFEYLVVA 247  
Db 207 LTGAQVVERGTFDIYIGSGQPLVIGNTNTLSTVPLKGPSPRMGIQWDRSGSTIDITSA 266  
QY 248 MNPSEDG-----SAASGTDASG----- 264  
Db 267 MTGGBIGLLTYRKEVLDPALNELGRVALVADQINRQQAQGDKNQDQGAALFNNINGA 326  
QY 265 -----LIMSGT-----MTFSSNGEL-----KNMTAFTGTSATKDLNAWQ-- 299  
Db 327 ALISQRTAQSGNSAGSNLDVTIKDTGKLTSDYQVTFITSATNYTVKRSDGTDMSFSST 386  
QY 300 ---PAPLVNGLPQFSANFVGAGIQLPLDIFGKSCQNNWACAPASAAA-----IGTDICK 351

Db 387 TTTTTPPVTDGF-----TLALNGALASAGDTFKVTPTRNAASSIQVLTDPKK 433  
QY 352 LPSMPPIOTSSGNSTARNSSSTRYSQDGYPOGDLVDV-----TITSEKLOG--KYNSN 405  
Db 434 IAAAGPL--TGVASNNSTG-----YTQPTLTDVVDIYNPASQAEQLNALKYSTP 481  
QY 406 QVVDYFNIPLARFTSEGLREGNNHYSATLDSGGPEFG---LPGTSNYKLSVNOLET 461  
Db 482 VKLVP-----GATTSQSQTYN--MVDKAGATIGSGVIVFGQANTLNKIGKWD 528  
QY 462 SNVDMRSREVMNMII--IQRFQMN 484  
Db 529 TGAPV--MDTVPINVKITFTVQT 550  
RESULT 35  
US-10-335-977-4779  
; Sequence 4779, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 427-7400  
; TELEFAX: (617) 742-4214  
; INFORMATION FOR SEQ ID NO: 4779:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 269 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...269  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4779:  
US-10-335-977-4779

Query Match 5.9%; Score 151.5; DB 15; Length 269;  
Best Local Similarity 25.9%; Pred. No. 0.00057;  
Matches 58; Conservative 39; Mismatches 90; Indels 37; Gaps 8;  
QY 6 FIGATG-MKCHSTGLGTVNNIANANTIGYKQQQWVFQD-----LFSODLAI 51  
Db 5 YTAATGATQFNRLDTSNNLANLNTNGPRDDAITGDFLRLYQYRQLPLEDTKAS 64

|    |    |     |                         |                        |                   |     |
|----|----|-----|-------------------------|------------------------|-------------------|-----|
| 52 | QY | 52  | GSTGSGNQAGMGAGVCSVTIT   | TTQGAFFPGNSVTDLAIGGK-  | PFQVLTLEKDVHYTSA  | 110 |
|    |    |     |                         |                        |                   |     |
|    | Db | 65  | AXYLNUNRNPUISEIYDTSRL   | ---GAFETGNPLDFALTSNLFY | ALQTNEGAYTFR      | 121 |
|    |    |     |                         |                        |                   |     |
|    | QY | 111 | GNFRQTQGFINDSGFTLMGSRIS | NN-----PNIKKETLEP      | QLQDFNDPTVAKSPA   | 162 |
|    |    |     |                         |                        |                   |     |
|    | Db | 122 | GHSVDKGFVLTVNGFKVL-     | SRSGLNKEGIMLMPNNAE     | TEVQNGGITFRD----- | 175 |
|    | QY | 163 | KTSTALNAVNLGSDTDTKQSEAN | PYFALLSEWKNGT          | PPPISTSS          | 206 |
|    |    |     |                         |                        |                   |     |
|    | Db | 176 | QIQAGALAVSFSPQNLKIKQNLV | -----TYQGEGVH          | QVSDS             | 214 |

RESULT 36

US-10-282-122A-64364  
; Sequence 64364, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

|     | Query Match           | 5.9%                                                    | Score 151.5;     | DB 15;      | Length 2204; |
|-----|-----------------------|---------------------------------------------------------|------------------|-------------|--------------|
|     | Best Local Similarity | 20.3%;                                                  | Pred. No. 0.012; |             |              |
|     | Matches 119;          | Conservative 64;                                        | Mismatches 220;  | Indels 185; | Gaps 24;     |
| QY  | 3                     | GSLEIGATGKMTSTGLGTVSNNIANVTIGYKQQOVFQDLFSODLAIAGTG----  | 55               |             |              |
| Ddb | 1458                  | ::     ::  : ::     : : :     :                         |                  |             |              |
| QY  | 56                    | SQGPNQAQMGAQVSGSVRTIFTQCAEPFGNSVDTLAIGKK-----GFQGV----- | 99               |             |              |

|    |      |                      |                 |                  |               |                   |              |             |            |             |         |      |     |
|----|------|----------------------|-----------------|------------------|---------------|-------------------|--------------|-------------|------------|-------------|---------|------|-----|
| Db | 1516 | NSGSFNTGLG - NTGSTNT | ----            | GLFNPGNVNTGVNNTG | INTGSI        | NTGSGFNTGSGT      | NTGSGF       | 1570        |            |             |         |      |     |
| Qy | 100  | TLED -               | ----            | KVHYTRAGNFRF -   | :             | TQDGF -           | ----         | 120         |            |             |         |      |     |
| Db | 1571 | NLGDHNTG             | SFGSDYNTGYFNAGD | YNTGVANTGVNTG    | AFISG         | SYNSNGFWRGDYQGLIC | 1630         |             |            |             |         |      |     |
| Qy | 121  | -----                | ----            | LANDPSGFTLMCSR   | SNPN -        | -----             | ----         | 147         |            |             |         |      |     |
| Db | 1631 | LSTTI                | PIPEI           | PRYDLS           | VPIDIPITGV    | ATTNFTIFCFQIRVLLG | PAALVNMEMIGP | 1690        |            |             |         |      |     |
| Qy | 148  | IQLDFNDPT            | VAKSPAKT        | STALANVNLGD      | STDKTQSEANPYF | FALLEBSWKNGTGPPI  | STEN         | 207         |            |             |         |      |     |
| Db | 1691 | ITIDV                | NQVIALDS        | FIQOTIS          | MSVGTGFG      | PIPIGISIGG        | PGF -        | -----       | 1736       |             |         |      |     |
| Qy | 208  | YSYAQ                | PRYVDQ          | QCN              | SHDITVFDG     | APSSTGSKT         | TEYLVAMP     | PEDSGSAAG - | TDSAGLL    | 266         |         |      |     |
| Db | 1737 | ---                  | TGPSG           | FFHTG            | AGHVSFG       | FGNFGAGN          | SGSGNF -     | -----       | -----      | 1782        |         |      |     |
| Qy | 267  | MSGTMT               | FSS -           | NGELK            | MNTAFTPTG     | SATKDLNA          | QWAPLVN      | LPQPSANFV   | GAGIO -    | ---PL       | 322     |      |     |
| Db | 1783 | NSGLL                | NFGLQ           | SLGANL           | GNTISGV       | YNTSLDL           | TPA -        | FSG         | GIANTGANL  | AGLFLONTGNL | 1841    |      |     |
| Qy | 323  | TLD                  | FGIKSQ          | NNWAG            | APASAA        | IGTDDIGKLP        | MMPIQTSS     | NGNSTARN    | SSSTRY -   | -----       | 377     |      |     |
| Db | 1842 | TLN                  | FVANQGL         | NAG -            | -----         | IGNLGS            | WNIGFVNTG    | SDNLGIGNL   | GLNLFNGVNI | 1890        |         |      |     |
| Qy | 378  | ---                  | SQDGPQ          | GDVDV            | TTITSE        | GKLOQ             | KYSN -       | SQVDF -     | -----      | YNTPLARE    | FTSED - | ---- | 422 |
| Db | 1891 | GGN                  | GIANTG          | TFDGL            | ANLGS         | YNTGLANL          | GDNDNLG      | FGNAGSYNL   | GFANFGSD   | NLGFANT     | 1950    |      |     |
| Qy | 423  | ----                 | GLRR            | BGNHYS -         | -----         | ATLDSG            | GPFG         | L -         | PGTSNYG    | 452         |         |      |     |
| Db | 1951 | GSYNT                | CFANTGN         | NTTGV            | LGTG          | NGOIGIG           | SLNSG        | SNNGI       | GLFN       | SGSGNIG     | 1998    |      |     |

RESULT 37

```

US-10-193-764-61
/ Sequence 61, Application US/10193764
/ Publication No. US20030133943A1
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M.
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michel H.
/ TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
/ TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
/ FILE REFERENCE: 1038-1239MIS
/ CURRENT APPLICATION NUMBER: US/10/193,764
/ CURRENT FILING DATE: 2002-07-12
/ PRIOR APPLICATION NUMBER: 09/167,568
/ PRIOR FILING DATE: 1998-10-07
/ NUMBER OF SEQ ID NOS: 91
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 61
/ LENGTH: 1180
/ TYPE: PRT
/ ORGANISM: Haemophilus influenzae
US-10-193-764-61

```

|                       |              |                                                               |                 |                      |
|-----------------------|--------------|---------------------------------------------------------------|-----------------|----------------------|
| Query Match           | 5.8%;        | Score 149.5;                                                  | DB 14;          | Length 1180;         |
| Best Local Similarity | 20.2%;       | Pred. No. 0.0072;                                             |                 |                      |
| Matches 113;          | Conservative | 88;                                                           | Mismatches 212; | Indels 147; Gaps 22; |
| QY                    | 30           | NTGYKQOVVFQDLFESQDLAISTGSGPNCAGCAQVGSVRIFQTQAFEPGNSVTDL       | 89              |                      |
| Db                    | 606          | NISGFNKAKIVAKD--SSNLTGNSDDSGNT-----SAKTV-----TFNNVKDS         | 647             |                      |
| QY                    | 90           | AIGSGKFFQVLTEDKHYVFRAGNFEFTQDGLNDPSGFTLGMGRISNNPNIKETLEPIQ    | 149             |                      |
| Db                    | 648          | KISADG-HKVTLNSKVK-TLSDNDNTEGSDNN-TGLTITAKDEVNNHIT-----        | 696             |                      |
| QY                    | 150          | LDNFNDTVAKSPA-----KTSALNAVNIIGDSTDKTQS-----EANPYALLESW---     | 195             |                      |
| Db                    | 697          | ---SHKTIVNVAANGIITTKTGTTINATAGNVEITATHTGTGQIGIESKPGSVITVAGGDT | 753             |                      |

QY 196 -----KGN-----GTPPISTNSYSYQPMRVYDQGN----- 222  
Db 754 LAVGNISGNVTVTNSGALTTLTLAGSTIKGTESITSS-----QSGNIGKISG 802  
QY 223 -----SHDITVYFDGAPSPSTGKTFEYLVANMPSDEGSAASG-----TDSAG 264  
Db 803 KTVNKRATNSLTQADSKIEATEG-----EANTVSKTSIIGGTISGGTVEVTATEGLTTQAG 859  
QY 265 LLMSGTWTSSNGELKNMTAFTPTG-----SATKDL-----NMQPAPLVN 305  
Db 860 STITGESVTTSSQSGNIGMISGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 919  
QY 306 GLPQFSANFVGAGIQTLDLDFGIKSQQNMWAGAPASAAAIGTDIGKLPSPMPTQSSGNS 365  
Db 920 GTISGNTVNTANTGDLTVEDAK-----IDATGGAATLTATSGKL-----TTKASS 966  
QY 366 TARGSSSTRYSQDGYPOQDL--VDVTTTSEGKLOGKYSNSQVDFVNIPLARFTSDG 423  
Db 967 SITSAQNVLNLSAKDGSIGGNINAANVTLTGALTTVKSSINANSGLVINAKDAELN 1026  
QY 424 LRREGNNHYSATLDSGGPEFGLPCTSNYKLSVNOLETNSVD--MSREVMNMIIQRFQM 482  
Db 1027 GEASGNTVNTANANGSGSVIATTSRVNITGDLITINGLNIISKNGINTVLL-KGVKI 1085  
QY 483 NKSXVTTADTLMQALELKR 502  
Db 1086 DVKIOPGIASVDEVIEAKR 1105

RESULT 38  
US-10-193-764-59  
; Sequence 59, Application US/10193764.  
; Publication No. US2003013943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US/10/193,764  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 59  
; LENGTH: 1188  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-59

Query Match 5.8%; Score 149.5; DB 14; Length 1188;  
Best Local Similarity 20.2%; Pred. No. 0.0073;  
Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;  
QY 30 NTIGYKQOVVFDLFSODLAIGSTGSGGNFQAGVGSVRTFTQGAPEGNSVTDL 89  
Db 614 NISGFNKAIVAKD--SSNLITGNSDDSGNT-----SAKTV-----TFNNVKDS 655  
QY 90 AIGKGFFQVTLKDVHYTRAGNFRFTQDGLNDPDSGFTLMGSRISNNPNIKKLEPIQ 149  
Db 656 KISADG-KHVTLNSKV-KTSLDNDNTEGSDNN-TGLTITAKDVEVNNIT----- 704  
QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTS-----EANPYFALLESW--- 195  
Db 705 ---SHKTVNVAANGGIITKTGTITINATAGNVEITATGSGTQGGIESKPGSVITVAGGDT 761  
QY 196 -----KGN-----GTPPISTNSYSYQPMRVYDQGN----- 222  
Db 762 LAVGNISGNVTVTNSGALTTLTLAGSTIKGTESITSS-----QSGNIGKISG 810

QY 223 -----SHDITVYFDGAPSPSTGKTFEYLVANMPSDEGSAASG-----TDSAG 264  
Db 811 KTVNKRATNSLTQADSKIEATEG-----EANTVSKTSIIGGTISGGTVEVTATEGLTTQAG 867  
QY 265 LLMSGTWTSSNGELKNMTAFTPTG-----SATKDL-----NMQPAPLVN 305  
Db 868 STITGESVTTSSQSGNIGMISGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 927  
QY 306 GLPQFSANFVGAGIQTLDLDFGIKSQQNMWAGAPASAAAIGTDIGKLPSPMPTQSSGNS 365  
Db 928 GTISGNTVNTANTGDLTVEDAK-----IDATGGAATLTATSGKL-----TTKASS 974  
QY 366 TARGSSSTRYSQDGYPOQDL--VDVTTTSEGKLOGKYSNSQVDFVNIPLARFTSDG 423  
Db 975 SITSAQNVLNLSAKDGSIGGNINAANVTLTGALTTVKSSINANSGLVINAKDAELN 1034  
QY 424 LRREGNNHYSATLDSGGPEFGLPCTSNYKLSVNOLETNSVD--MSREVMNMIIQRFQM 482  
Db 1035 GEASGNTVNTANANGSGSVIATTSRVNITGDLITINGLNIISKNGINTVLL-KGVKI 1093  
QY 483 NKSXVTTADTLMQALELKR 502  
Db 1094 DVKIOPGIASVDEVIEAKR 1113

RESULT 39  
US-10-335-977-4780  
; Sequence 4780, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 4780:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 273 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...273  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4780:





Db 171 LGFDKSSSTANPPFGMLEAWDGTATPPLGNAAGYTQTMRVYDAAGDTHLTIKFDNA 230  
 QY 234 PSTSGSKTEYLVANPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293  
 Db 231 TTAGGKRYVEFLVMTTPPEDGGAAGSAGAGLLMSGTLQFGSSGQLQDIMAFTPSGGDPK 290  
 QY 294 DLNAPQAPL-VNGLPQFSANFVGAGIQPLTLDGFIKQQNNWAGAPASAAAIGTDIGKL 352  
 Db 291 DLNANVPATLDASGRPQFNATFAOQGAQTVGLDLGIT--TANNAPASAAAGVGLDPTLL 348  
 QY 353 PSMWPIQTSSGNSTARNSSSTRYSODGYPOGDLVDVTITSEGLKQKYSNSQVDFYN 412  
 Db 349 GGATTPKLAATSTTAAYKSSSTYKQDGYAAGVLMNLEAVTDGIVSGKYSNGOSQLFR 408  
 QY 413 IPLARFTSEDLGRREGNNHYSATLDSGGPEFGLPCTSNYKLSVNQLETSNVDMREMN 472  
 Db 409 VPIFTSEDLGRSEGNNHYSATTESGAQEGKADTENYKLGKMGKSLQSNVDMAREMN 468  
 QY 473 MIIORGQFOMNSKSVTTADTMLOKALELKR 502  
 Db 469 MIVTORGQFOTNSKVTTADTMIOKALELKR 498

## RESULT 2

AAS94790 ID AAS94790 PRELIMINARY; PRT; 498 AA.  
 AC AAS94790; Q72C41; AC AAS94790; PRT; 498 AA.  
 DT 27-APR-2004 (TREMELrel. 27, Created)  
 DT 27-APR-2004 (TREMELrel. 27, Last sequence update)  
 DT 11-MAY-2004 (TREMELrel. 27, Last annotation update)  
 DE Flagella basal body rod domain protein.  
 GN DVU0307.  
 OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;  
 OC Desulfovibrionaceae; Desulfovibrio.  
 OX NCBI\_TaxID=882;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed15077118; DOI=10.1038/nbt959;  
 RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,  
 RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,  
 RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,  
 RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,  
 RA Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,  
 RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,  
 RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;  
 RT "The genome sequence of the anaerobic, sulfate-reducing bacterium  
 Desulfovibrio vulgaris Hildenborough.";  
 RL Nat. Biotechnol. 22:554-559(2004).  
 DR EMBL; AE017310; AAS94790.1; -.  
 DR TIGR; DVU0307; -.  
 KW Flagella.  
 SQ SEQUENCE 498 AA; 51954 MW; CDD6B383774E8807 CRC64;  
 Query Match 47.8%; Score 1235; DB 2; Length 498;  
 Best Local Similarity 49.6%; Pred. No. 5.1e-67;  
 Matches 253; Conservative 74; Mismatches 163; Indels 20; Gaps 5;

QY 1 MMGSLFICATGCMKTHSTGLGTVSNANANTIGYKQQVVFQDLFSODLAIGSTGSGGN 60  
 Db 1 MMHGLYTGATGMRTQAEGMNVGNNLANVTIGKQSMMLYQDLSQTEPTGSAVYSGIS 60  
 QY 61 QAGMGAQVGSVRTFTQGAPEPGNSVTDLAIGKGFQVLTEDKHVYTRAGNFRFTQDGF 120  
 Db 61 QVGLGARGVDVRLRSQALLAGSDITDFALSGKGFQVTSQGDTHYTRAGNFRFNKQ 120  
 QY 121 LNDPSGTLMGSRISNNPNIKETLEPIQLDFNDPTVAKS-----PAKTSTALNAVYN 173  
 Db 121 LVDPNGLNMGHAITG-----ETEGPLAA-----VTLVKDADGRLSNPAKATGMSVFN 170  
 QY 174 LGDSTDKTQSEANPYFALLESWKNGTTPPTISNYSYAQPMRVYDQGNSHDITVYFDCA 233

Db 171 LGFDKSSSTANPPFGMLEAWDGTATPPLGNAAGYTQTMRVYDAAGDTHLTIKFDNA 230  
 QY 234 PSTSGSKTEYLVANPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293  
 Db 231 TTAGGKRYVEFLVMTTPPEDGGAAGSAGAGLLMSGTLQFGSSGQLQDIMAFTPSGGDPK 290  
 QY 294 DLNAPQAPL-VNGLPQFSANFVGAGIQPLTLDGFIKQQNNWAGAPASAAAIGTDIGKL 352  
 Db 291 DLNANVPATLDASGRPQFNATFAOQGAQTVGLDLGIT--TANNAPASAAAGVGLDPTLL 348  
 QY 353 PSMWPIQTSSGNSTARNSSSTRYSODGYPOGDLVDVTITSEGLKQKYSNSQVDFYN 412  
 Db 349 GGATTPKLAATSTTAAYKSSSTYKQDGYAAGVLMNLEAVTDGIVSGKYSNGOSQLFR 408  
 QY 413 IPLARFTSEDLGRREGNNHYSATLDSGGPEFGLPCTSNYKLSVNQLETSNVDMREMN 472  
 Db 409 VPIFTSEDLGRSEGNNHYSATTESGAQEGKADTENYKLGKMGKSLQSNVDMAREMN 468  
 QY 473 MIIORGQFOMNSKSVTTADTMLOKALELKR 502  
 Db 469 MIVTORGQFOTNSKVTTADTMIOKALELKR 498

## RESULT 3

Q72C41 ID Q72C41 PRELIMINARY; PRT; 564 AA.  
 AC Q72C41; Q72C41; AC Q72C41; PRT; 564 AA.  
 DT 05-JUL-2004 (TREMELrel. 27, Created)  
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)  
 DE Flagellar hook protein Flg3.  
 GN Name=flgB; Order=locusNames=DVU1443;  
 OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;  
 OC Desulfovibrionaceae; Desulfovibrio.  
 OX NCBI\_TaxID=882;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed15077118; DOI=10.1038/nbt959;  
 RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,  
 RA Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,  
 RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,  
 RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,  
 RA Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,  
 RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,  
 RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;  
 RT "The genome sequence of the anaerobic, sulfate-reducing bacterium  
 Desulfovibrio vulgaris Hildenborough.";  
 RL Nat. Biotechnol. 22:554-559(2004).  
 DR EMBL; AE017314; AAS95921.1; -.  
 DR TIGR; DVU1443; -.  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR01444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 KW Complete proteome; Flagellum.  
 SQ SEQUENCE 564 AA; 60342 MW; 50DC0B4F6D7A4DBE CRC64;

Query Match 37.2%; Score 961; DB 2; Length 564;  
 Best Local Similarity 39.0%; Pred. No. 2.9e-50;  
 Matches 223; Conservative 85; Mismatches 184; Indels 80; Gaps 17;  
 QY 1 MMGSLFICATGCMKTHSTGLGTVSNANANTIGYKQQVVFQDLFSODLAIGSTGSGGN 60  
 Db 3 LTASWTVGSGLLAGGERMNVGNANVNVTVGKSGRMDFEFLNOD-----TVSAAGVT 58  
 QY 61 QAGMGAQVGSVRTFTQGAPEPGNSVTDLAIGKGFQVLTEDKHVYTRAGNFRFTQD 118  
 Db 59 QVGRGVSIGALFGDYSQACAFQTTNESTDLATQGRGFSFVKPGTGYTTRAGNFRPDAD 118

```

QY 119 GFLNDPSGFTLMGSRISNNPN-----IKKTELEPIQLDFNDPTVAKSPAK 163
DB 119 GYLVDPHGVVLOGWAIERSONSLTSSAVTATSTTSKIKSGV-PVDIKL-DGFTAEPOHT 176
QY 164 TSTALNAVNLGDSUTKQSEANPYFALLESKWG-----NGT-PPISTSNYSYAPQMRVVD 218
DB 177 QNITLNVNLDSPGNDKSSSTNPNPFSLFETWNGONPLTGTQPALAQSAFAYQSTIKVYD 236
QY 219 OQGNSHDITVYFDGA-PSS-----TGSKTFEYLVAMPSED-----GSAASGTDASGLLM 267
DB 237 EAGTAHTLTIVYFDQVDPDSVTNEPNRQKWEYIVTMDPAEDKRVIACTAMNTTAAAGLLM 296
QY 268 SGTMTFSSNGELKNTAFTPTGSAT-----KDLNMQAPL-VNG 306
DB 297 TGLTFDTGQLVDQTATFTPGQYDTTPPNNEPTNPGPPVTPPALVNVQPMQSSNG 356
QY 307 LPQSFANFVG-----AGIOLPLDFGIKSQQ--NNWAGAPASAAAIGTDIGKLP 353
DB 357 LPMVAVNSGLTDSVVGSPTAQNFMELDLGLASTNATPTWSTP-NAAAIGTDASLLP 415
QY 354 SMPTQTSNGNSTARNSSSTRYSODGYPOGDLVDVTTITSEGLKQKYSNSQVVDPYNI 413
DB 416 GLTSSQRPQSATTSVAGSSSTQFQKQGYTFGLQNTIVDRDGMQKYSNGVTLDIYQV 475
QY 414 PLARFSDGLRRGNHYSATLDSGGPEFGLPTSN---YGLKSVNQLTNSVNDMSREM 470
DB 476 TLVDETSKONLAREGGLNFGSTRDSGP---LPGPANSNGLGAISSNSLSQSNVDLAREF 532
QY 471 VNMIIQRFQFQNSKSVTTADTMLOKALELKR 502
DB 533 VEMITQRFQFQNSKSVITTTDMLEVVVNMKR 564

RESULT 4
ID AAS95921 PRELIMINARY; PRT; 564 AA.
AC AAS95921;
DT 27-APR-2004 (TremBLrel. 27, Created)
DT 27-APR-2004 (TremBLrel. 27, Last sequence update)
DE 11-MAY-2004 (TremBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN FLGE OR DVU1443.
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobibrionales;
OC Desulfobibrionaceae; Desulfobivrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RL Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; A2017314; AAS95921.1; --
DR TIGR; DVU1443; --
KW Flagella.
SQ SEQUENCE 564 AA; 60542 MW; 50DC0B4F6D7A4DBE CRC64;

Query Match 37.2%; Score 961; DB 2; Length 564;
Best Local Similarity 39.0%; Pred. No. 2.9e-50;
Matches 223; Conservative 85; Mismatches 184; Indels 80; Gaps 17;

QY 1 MMGSLPIGATGKHTSTGLGTVSNNIANTIGKQQVQFDLPFSDLAIGSTGSGPN 60
DB 3 LTASMTGVSGLLAHGERANVLGNINANTVGVFGKSRMDFEDFLNQD----TYSAGV 58

```

```

QY 61 QAGMGAQVGSVRTITFQGAFFPGNSVTDLAIGKGKFFQVTL--DKVHYTRAGNFRFTQD 118
DB 59 QVGRGVSGIAGFDGYSQGAFQTTNSTDLAIQGRFPFVKPGTGTEDTYTRAGNFRFDAD 118
QY 119 GFLNDPSGFTLMGSRISNNPN-----IKKTELEPIQLDFNDPTVAKSPAK 163
DB 119 GYLVDPHGVVLOGWAIERSONSLTSSAVTATSTTSKIKSGV-PVDIKL-DGFTAEPOHT 176
QY 164 TSTALNAVNLGDSUTKQSEANPYFALLESKWG-----NGT-PPISTSNYSYAPQMRVVD 218
DB 177 QNITLNVNLDSPGNDKSSSTNPNPFSLFETWNGONPLTGTQPALAQSAFAYQSTIKVYD 236
QY 219 OQGNSHDITVYFDGA-PSS-----TGSKTFEYLVAMPSED-----GSAASGTDASGLLM 267
DB 237 EAGTAHTLTIVYFDQVDPDSVTNEPNRQKWEYIVTMDPAEDKRVIACTAMNTTAAAGLLM 296
QY 268 SGTMTFSSNGELKNTAFTPTGSAT-----KDLNMQAPL-VNG 306
DB 297 TGLTFDTGQLVDQTATFTPGQYDTTPPNNEPTNPGPPVTPPALVNVQPMQSSNG 356
QY 307 LPQSFANFVG-----AGIOLPLDFGIKSQQ--NNWAGAPASAAAIGTDIGKLP 353
DB 357 LPMVAVNSGLTDSVVGSPTAQNFMELDLGLASTNATPTWSTP-NAAAIGTDASLLP 415
QY 354 SMPTQTSNGNSTARNSSSTRYSODGYPOGDLVDVTTITSEGLKQKYSNSQVVDPYNI 413
DB 416 GLTSSQRPQSATTSVAGSSSTQFQKQGYTFGLQNTIVDRDGMQKYSNGVTLDIYQV 475
QY 414 PLARFSDGLRRGNHYSATLDSGGPEFGLPTSN---YGLKSVNQLTNSVNDMSREM 470
DB 476 TLVDETSKONLAREGGLNFGSTRDSGP---LPGPANSNGLGAISSNSLSQSNVDLAREF 532
QY 471 VNMIIQRFQFQNSKSVTTADTMLOKALELKR 502
DB 533 VEMITQRFQFQNSKSVITTTDMLEVVVNMKR 564

RESULT 5
QY 61 QAGMGAQVGSVRTITFQGAFFPGNSVTDLAIGKGKFFQVTL--DKVHYTRAGNFRFTQD 118
DB 59 QVGRGVSGIAGFDGYSQGAFQTTNSTDLAIQGRFPFVKPGTGTEDTYTRAGNFRFDAD 118
QY 119 GFLNDPSGFTLMGSRISNNPN-----IKKTELEPIQLDFNDPTVAKSPAK 163
DB 119 GYLVDPHGVVLOGWAIERSONSLTSSAVTATSTTSKIKSGV-PVDIKL-DGFTAEPOHT 176
QY 164 TSTALNAVNLGDSUTKQSEANPYFALLESKWG-----NGT-PPISTSNYSYAPQMRVVD 218
DB 177 QNITLNVNLDSPGNDKSSSTNPNPFSLFETWNGONPLTGTQPALAQSAFAYQSTIKVYD 236
QY 219 OQGNSHDITVYFDGA-PSS-----TGSKTFEYLVAMPSED-----GSAASGTDASGLLM 267
DB 237 EAGTAHTLTIVYFDQVDPDSVTNEPNRQKWEYIVTMDPAEDKRVIACTAMNTTAAAGLLM 296
QY 268 SGTMTFSSNGELKNTAFTPTGSAT-----KDLNMQAPL-VNG 306
DB 297 TGLTFDTGQLVDQTATFTPGQYDTTPPNNEPTNPGPPVTPPALVNVQPMQSSNG 356
QY 307 LPQSFANFVG-----AGIOLPLDFGIKSQQ--NNWAGAPASAAAIGTDIGKLP 353
DB 357 LPMVAVNSGLTDSVVGSPTAQNFMELDLGLASTNATPTWSTP-NAAAIGTDASLLP 415
QY 354 SMPTQTSNGNSTARNSSSTRYSODGYPOGDLVDVTTITSEGLKQKYSNSQVVDPYNI 413
DB 416 GLTSSQRPQSATTSVAGSSSTQFQKQGYTFGLQNTIVDRDGMQKYSNGVTLDIYQV 475
QY 414 PLARFSDGLRRGNHYSATLDSGGPEFGLPTSN---YGLKSVNQLTNSVNDMSREM 470
DB 476 TLVDETSKONLAREGGLNFGSTRDSGP---LPGPANSNGLGAISSNSLSQSNVDLAREF 532
QY 471 VNMIIQRFQFQNSKSVTTADTMLOKALELKR 502
DB 533 VEMITQRFQFQNSKSVITTTDMLEVVVNMKR 564

PRT; 419 AA.
PRELIMINARY;
ID Q74G30
AC Q74G30;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN Name=flgE; ORFNames=GSU0419;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PCA / ATCC 51573;
RA Heidelberg J.F., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
RA Davidsson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
RA Weidman J.P., Khouri H.M., Feldblyum T.V., Utterback T.R.,
RA Van Aken S.E., Lovley D.R., Fraser C.M.;
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RL Science 302:1967-1969(2003).
DR EMBL; AEO17180; AAR33751.1; --
DR TIGR; GSU0419; --
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb rod.
DR InterPro; IPR002371; Flag_hookAPI.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb rod; 1.

```

```

DR PRINTS; PRO1005; FLGHOOKAPI.
DR PROSITE; PS00588; FLAGELLA_BB_OD; 1.
KW Flagellum.
SQ SEQUENCE 419 AA; 44051 MW; 3148770150C5C6EF CRC64;

Query Match 26.8%; Score 693.5; DB 2; Length 419;
Best Local Similarity 34.9%; Pred. No. 3.7e-34;
Matches 175; Conservative 70; Mismatches 166; Indels 91; Gaps 13;

QY 4 SLFIGATGKTHSTGLGTVSNNIANANTIGYKQQQVWFQDLFSQDLAIGTSGTSGQPNQAG 63
DB 6 ALYTGISGLNANGEAMSVIGNNISNNTIGFGQRMFLFSDVLSSTISGGS-----QIG 58

QY 64 MGAQVGSVRTITQGAPEFGNSVTDLAIGGKGFQVTTLEDKVHYTRAGNFRFTQDGFIND 123
DB 59 RGVOIQTVENQTFQSFESTESGTDLAIOGDSFFVQNTSGRYYTRAGAFSFKDKTLVN 118

QY 124 PSQFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNLGDSGTDKTS 183
DB 119 PEGYQWVGIIIPSSGLSDGVLPIDL-----TNFATTPKQTSVKFVNVL-DSTQTTPT 173

QY 184 EANPYFALLESKWNGTTPPISTNSYSAQPMRVDDQGNSHDITVYFDGAPSGTSGKTFE 243
DB 174 LA-----W--DPANPVATSNYSTS--LSVYDSQGNHTATVYF-----RKTADNWD 216

QY 244 YLVAMNPSEDGSAAGTDSAGLLMSGTWTFSNGELKMTAFTPTGSAKDNLNAWQAPL 303
DB 217 WHVILPDAAAGTGGSTTTP-----IDGTLTFDATGAL-----TAQTPLAGAAQNI----- 261

QY 304 VNGLPQFSANFVGAGI-----QPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLPMMPIQT 360
DB 262 -----TFAG-GVTAPQIFFDLGV----- 279

QY 361 SSGNSTARNGSSSTRYSQDGYPGDLVDVITITSEGLKQKYSNVVDVFNIPARFTS 420
DB 280 --GATTQVASSVSSQTQDGYQGLTKVITDDKGYVNGVYSGQKLYQVALAKFSS 337

QY 421 EDGLRREGNNHYATLDSGGPEFGPLGTSNYGKLSVNOLETNSVDMREVMNMIIOQGF 480
DB 338 TAGLSKAGGTLFEETLESQGLFSDASTPGVGKILANSLEQSNVDMMAQFVMMITTOQGY 397

QY 481 QMNSKSVTTADTMLQKALELKR 502
DB 398 SANSKTIITADEMLQEVLSLKR 419

RESULT 6
AAR33751 PRELIMINARY; PRT; 419 AA.
AC AAR33751;
DT 02-MAR-2004 (TReMBLrel. 27, Created)
DT 02-MAR-2004 (TReMBLrel. 27, Last sequence update)
DE Flagellar hook protein FlgE.
GN FLGE OR GSU0419
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA Heidelberg J.F., Wu D., Wu M., Ward N.H., Beanan M.J., Dodson R.J.,
RA Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
RA Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
RA Weidman J., Khouri H.M., Feldblyum T.V., Utterback T.R.,
RA Van Aken S.E., Lowley D.R., Fraser C.M.;
RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface
RT environments.";
RL Science 302:11967-1969 (2003).

DR EMBL; AE017208; AAR33751.1; -.
DR TIGR; GSU0419; -.
KW Flagella.
SQ SEQUENCE 419 AA; 44051 MW; 3148770150C5C6EF CRC64;

Query Match 26.8%; Score 693.5; DB 2; Length 419;
Best Local Similarity 34.9%; Pred. No. 3.7e-34;
Matches 175; Conservative 70; Mismatches 166; Indels 91; Gaps 13;

QY 4 SLFIGATGKTHSTGLGTVSNNIANANTIGYKQQQVWFQDLFSQDLAIGTSGTSGQPNQAG 63
DB 6 ALYTGISGLNANGEAMSVIGNNISNNTIGFGQRMFLFSDVLSSTISGGS-----QIG 58

QY 64 MGAQVGSVRTITQGAPEFGNSVTDLAIGGKGFQVTTLEDKVHYTRAGNFRFTQDGFIND 123
DB 59 RGVOIQTVENQTFQSFESTESGTDLAIOGDSFFVQNTSGRYYTRAGAFSFKDKTLVN 118

QY 124 PSQFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNLGDSGTDKTS 183
DB 119 PEGYQWVGIIIPSSGLSDGVLPIDL-----TNFATTPKQTSVKFVNVL-DSTQTTPT 173

QY 184 EANPYFALLESKWNGTTPPISTNSYSAQPMRVDDQGNSHDITVYFDGAPSGTSGKTFE 243
DB 174 LA-----W--DPANPVATSNYSTS--LSVYDSQGNHTATVYF-----RKTADNWD 216

QY 244 YLVAMNPSEDGSAAGTDSAGLLMSGTWTFSNGELKMTAFTPTGSAKDNLNAWQAPL 303
DB 217 WHVILPDAAAGTGGSTTTP-----IDGTLTFDATGAL-----TAQTPLAGAAQNI----- 261

QY 304 VNGLPQFSANFVGAGI-----QPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLPMMPIQT 360
DB 262 -----TFAG-GVTAPQIFFDLGV----- 279

QY 361 SSGNSTARNGSSSTRYSQDGYPGDLVDVITITSEGLKQKYSNVVDVFNIPARFTS 420
DB 280 --GATTQVASSVSSQTQDGYQGLTKVITDDKGYVNGVYSGQKLYQVALAKFSS 337

QY 421 EDGLRREGNNHYATLDSGGPEFGPLGTSNYGKLSVNOLETNSVDMREVMNMIIOQGF 480
DB 338 TAGLSKAGGTLFEETLESQGLFSDASTPGVGKILANSLEQSNVDMMAQFVMMITTOQGY 397

QY 481 QMNSKSVTTADTMLQKALELKR 502
DB 398 SANSKTIITADEMLQEVLSLKR 419

RESULT 7
Q9ROB6 PRELIMINARY; PRT; 463 AA.
AC Q9ROB6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN Name=flgE; ORFNames=IDS2768;
OS Treponema denticola ATCC 35405.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=243275;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35405;
RA Seshadri R., Myers G.S., Tettelin H., Eisen J.A., Heidelberg J.F.,
RA Dodson R.J., Davidson T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
RA Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F.,
RA Durkin A.S., Daugherty S.C., Shetty J., Shvartsbeyn A.,
RA Gebregeorgis E., Geer K., Tsegaye G., Malek J.A., Ayodeji B.,
RA Shatsman S., McLeod M.P., Majd D., Howell J.K., Pal S., Amin A.,
RA Vaishath P., McNeill T.Z., Weinstock G.M., Norris S.J., Fraser C.M.,
RA Paulsen I.T.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017255; AAI3285.1; -.
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.

```

DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0001539; P: ciliary/flagellar motility; IEA.  
 DR InterPro: IPR010930; DUF1078.  
 DR InterPro: IPR011491; FlaeB.  
 DR InterPro: IPR001444; Flag\_bb\_rod.  
 DR Pfam: PF06429; DUF1078; 1.  
 DR Pfam: PF07559; FlaeB; 1.  
 DR Pfam: PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE: PS00588; FLAGELLA\_BB\_ROD; 1.  
 KW Flagellum.  
 SQ SEQUENCE 463 AA; 49574 MW; E872E4C726F6C733 CRC64;

Query Match 23.0%; Score 595.5; DB 2; Length 463;  
 Best Local Similarity 32.2%; Pred. No. 3.9e-28;  
 Matches 166; Conservative 73; Mismatches 210; Indels 67; Gaps 12;

QY 1 MNGSLFICATGKMTSTGLTVSNNIANANTIGYKQOVVFDLPSDPLA-----IGSTG 55  
 DB 1 MNRSLFSGVTGQNHQTRMDVIGNVANVTGFKRGVNFQDLISQLSGAARPTBELG 60  
 QY 56 SOGPNQAGCAQVGSVRTIFTOGAFEPGNSVTDLAIGKGFQVTLDEKVVHYTRAGNFR 115  
 DB 61 GYNPKEVGLGMVVASIDIFTQALQTTGVNTDLAIQNGFFILKDGKFTYTRAGAFGI 120  
 QY 116 TDGFLNDPSGFTLMSRISNNPNIKETLEPIQL-----DFNDPTVAKSPAKTSTALNA 170  
 DB 121 DKEGTLVNPAN---GMRVQGMMAEEAEGFRINTSGQTELDNIPIGQKLDKATTSVNY 176  
 QY 171 VNLGSDTDKTOSEANPYFALESKNGKNGTPPISSTNSYVAQPMRVYDQCGNSHDITVYF 230  
 DB 177 ACNLDKRLPELPEGANRAQILESTW-----STEFKYVDSFGAEHELQIDF 221  
 QY 231 DGAPSTGSKTFEYLIVAMPNPSDEGSAAS-----GTDSAGLLMSGTMTSSNGELKNMFTAFT 286  
 DB 222 ARVPEVNA--WRATVNVDPNTADATATRVGIGT--TDGVQNSFIVRFDDNNGHLASVT--D 276  
 QY 287 PTGSATKDLNAQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQQNMWAGAPASAAAIG 346  
 DB 277 TAGNVT-----SPAGQVLVQISYVNVGANPDE-----AGAP--TRHTFD 313  
 QY 347 TDIGLPSMMPITQSSGNSTARGSSSTRYSQDGPQGLVDVTTITSEGLQKYSNSQ 406  
 DB 314 VNLGE-----IGTSKNITTFQSDKSTTKAYEQDGYTLGYLENFRIDQSGIITGVYSGV 367  
 QY 407 VVDFYNIPLARTSEDGLRRENNHYSATLDSGGPEFGLPCTSNYKLSVNLQLETSNVD 466  
 DB 368 ROEIGQIANAGFANQGLEKAGQNTYVQSNNGIANVSTSGTVGKGYFIGTLEMSNVDL 427  
 QY 467 SEEMVNMIIQRFQMSKSVTTADTMLQKALELKR 502  
 DB 428 TDQFVDMIVTQKGFQAGAKTIQTSDTMLETVNLKR 463

## RESULT 8

AAS13285  
 ID AAS13285 PRELIMINARY; PRT; 463 AA.  
 AC AAS13285;  
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Flagellar hook protein FlgE.  
 DE FLGE OR TDE2768.  
 GN Treponema denticola.  
 OS Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OC NCBI\_TaxID=158;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35405 / DSM 14222;  
 RX PubMed=15064399;  
 RA Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,  
 RA Dodson R.J., Davidlsen T.M., DeBoy R.T., Fouts D.E., Haft D.H.,  
 RA Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J., Durkin S.A.,  
 RA Daugherty S.C., Shetty J., Shvartsbeyn A., Gebregeorgis E., Gear K.,

RA Tsegaye G., Malek J.A., Ayodeji B., Shatsman S., McLeod M.P.,  
 RA Smajs D., Howell J.K., Pal S., Amin A., Vashisth P., McNeill T.Z.,  
 RA Xiang Q., Sodergren E., Baca E., Weinstein G.M., Norris S.J.,  
 RA Fraser C.M., Paulsen I.T.;  
 RT "Comparison of the genome of the oral pathogen Treponema denticola  
 RT with other spirochete genomes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).  
 DR EMBL: AB017255; AAS13285.1; -.  
 DR TIGR: TDE2768; -.  
 KW Flagella.  
 SQ SEQUENCE 463 AA; 49574 MW; E872E4C726F6C733 CRC64;

Query Match 23.0%; Score 595.5; DB 2; Length 463;  
 Best Local Similarity 32.2%; Pred. No. 3.9e-28;  
 Matches 166; Conservative 73; Mismatches 210; Indels 67; Gaps 12;

QY 1 MNGSLFICATGKMTSTGLTVSNNIANANTIGYKQOVVFDLPSDPLA-----IGSTG 55  
 DB 1 MNRSLFSGVTGQNHQTRMDVIGNVANVTGFKRGVNFQDLISQLSGAARPTBELG 60  
 QY 56 SOGPNQAGCAQVGSVRTIFTOGAFEPGNSVTDLAIGKGFQVTLDEKVVHYTRAGNFR 115  
 DB 61 GYNPKEVGLGMVVASIDIFTQALQTTGVNTDLAIQNGFFILKDGKFTYTRAGAFGI 120  
 QY 116 TDGFLNDPSGFTLMSRISNNPNIKETLEPIQL-----DFNDPTVAKSPAKTSTALNA 170  
 DB 121 DKEGTLVNPAN---GMRVQGMMAEEAEGFRINTSGQTELDNIPIGQKLDKATTSVNY 176  
 QY 171 VNLGSDTDKTOSEANPYFALESKNGKNGTPPISSTNSYVAQPMRVYDQCGNSHDITVYF 230  
 DB 177 ACNLDKRLPELPEGANRAQILESTW-----STEFKYVDSFGAEHELQIDF 221  
 QY 231 DGAPSTGSKTFEYLIVAMPNPSDEGSAAS-----GTDSAGLLMSGTMTSSNGELKNMFTAFT 286  
 DB 222 ARVPEVNA--WRATVNVDPNTADATATRVGIGT--TDGVQNSFIVRFDDNNGHLASVT--D 276  
 QY 287 PTGSATKDLNAQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQQNMWAGAPASAAAIG 346  
 DB 277 TAGNVT-----SPAGQVLVQISYVNVGANPDE-----AGAP--TRHTFD 313  
 QY 347 TDIGLPSMMPITQSSGNSTARGSSSTRYSQDGPQGLVDVTTITSEGLQKYSNSQ 406  
 DB 314 VNLGE-----IGTSKNITTFQSDKSTTKAYEQDGYTLGYLENFRIDQSGIITGVYSGV 367  
 QY 407 VVDFYNIPLARTSEDGLRRENNHYSATLDSGGPEFGLPCTSNYKLSVNLQLETSNVD 466  
 DB 368 ROEIGQIANAGFANQGLEKAGQNTYVQSNNGIANVSTSGTVGKGYFIGTLEMSNVDL 427  
 QY 467 SEEMVNMIIQRFQMSKSVTTADTMLQKALELKR 502  
 DB 428 TDQFVDMIVTQKGFQAGAKTIQTSDTMLETVNLKR 463

## RESULT 9

Q6MEY9  
 ID Q6MEY9 PRELIMINARY; PRT; 427 AA.  
 AC Q6MEY9;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE FlgE protein.  
 GN Name=flgE; OrderedLocusNames=Bd3395;  
 OS Bdellovibrio bacteriovorus.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;  
 OC Bdellovibrionaceae; Bdellovibrii.  
 OX NCBI\_TaxID=959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;  
 RX PubMed=14752164;  
 RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,  
 RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,  
 RA Sockett R.E., Schuster S.C.;

RT "A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a  
 RL genomic perspective.",  
 DR Science 303:689-692(2004).  
 DR EMBL; BX842655; CAE78193.1; -.  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR011491; Flae.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF07559; Flae; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 427 AA; 45943 MW; C74DA82F2E0FC003 CRC64;

Query Match 22.4%; Score 579.5; DB 2; Length 427;  
 Best Local Similarity 30.6%; Pred. No. 3.3e-27;  
 Matches 160; Conservative 73; Mismatches 171; Indels 119; Gaps 16;  
 QY 1 MMGSLFPGATGKTHSTGLTGVSNNTANANTIGYKQOVVFDLPSODLAIGSTGSGQPN 60  
 DB 3 ILSSLYTGVSGMTAQGEALGVIGDNIANANTIGFKASRAEFQDIISKNL-----KGIVG 58  
 QY 61 QAGMGAGVSVRTIFQGAFAFGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGF 120  
 DB 59 QIGRGVKGAVNPILSQGNIDATEKVTDLAISGDGYFKVKGSDGESYTRDGSFHFDRGY 118  
 QY 121 L--NDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDST 178  
 DB 119 LVTND-----NQRVQGFSTDEKGNIVNKMTDIKPPR-ALIPAKATKELKLDNLDSRM 170  
 QY 179 DKTQSEANPYFALLSSWKNGNTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSTG 238  
 DB 171 EPTK-----KFPDADPYSTSHYSTG--VEMYDSQGNKHLVSFFP-----NKVN 211  
 QY 239 SKTFEY--LVAMNPSDEGSAAGTDSAGL--LMSGTMTFSSNGELKNM-----TAFPTGS 290  
 DB 212 DREWEFKGLV-----DGKEITGEGEKMSVAAGKLMFTVDGKLDSEQTSTNPNFKGG 265  
 QY 291 ATKDLNMQAPLVLNGLPQFSANFVGAGIQLPTLDFG--IKSQQNMWAGAPASAAAI 348  
 DB 266 ALQD-----QQVKNLFGDAIK-----D 282  
 QY 349 IGKLPSPMPQIOTSSGNSTARNSSSTRYS-----QDGYPGDLVDVTTITSEGLK 399  
 DB 283 GKK-----GLDGTQYKGNKSLISWHQDGAAGTITGLSFNDEGTLT 324  
 QY 400 GKYSNQVVDVFNPIPLAFTSEDLRREGNNHYSATLDSGGPEPGLPSTNYGKLSVNQL 459  
 DB 325 AVYSNGQANDLAQIALAKFENPEALFKVGNRLKESRDSGTASVAGAPGAAGRGKLF 384  
 QY 460 ETSNVDMREVMNMIIOQGFQMSKSVTTADTLMQKALELKR 502  
 DB 385 ERSTVDLATEFVNMIQORGFQANAKITTTTDELLNEVIQLR 427

## RESULT 10

CAE78193  
 ID CAE78193 PRELIMINARY; PRT; 427 AA.  
 AC CAE78193;  
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE FlgE protein  
 GN FlgE OR B23395.  
 OS *Bdellovibrio bacteriovorus*.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; *Bdellovibrionales*;  
 OC *Bdellovibrionaceae*; *Bdellovibrio*.  
 OX NCBI\_TaxID=959;  
 RN [1]  
 RL SEQUENCE FROM N.A.  
 RP STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;  
 RX PubMed=14752164;  
 RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,

RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,  
 RA Sockett R.E., Schuster S.C.;  
 RT "A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a  
 RL genomic perspective.",  
 DR Science 303:689-692(2004).  
 DR EMBL; BX842655; CAE78193.1; -.  
 SQ SEQUENCE 427 AA; 45943 MW; C74DA82F2E0FC003 CRC64;

Query Match 22.4%; Score 579.5; DB 2; Length 427;  
 Best Local Similarity 30.6%; Pred. No. 3.3e-27;  
 Matches 160; Conservative 73; Mismatches 171; Indels 119; Gaps 16;  
 QY 1 MMGSLFPGATGKTHSTGLTGVSNNTANANTIGYKQOVVFDLPSODLAIGSTGSGQPN 60  
 DB 3 ILSSLYTGVSGMTAQGEALGVIGDNIANANTIGFKASRAEFQDIISKNL-----KGIVG 58  
 QY 61 QAGMGAGVSVRTIFQGAFAFGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGF 120  
 DB 59 QIGRGVKGAVNPILSQGNIDATEKVTDLAISGDGYFKVKGSDGESYTRDGSFHFDRGY 118  
 QY 121 L--NDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDST 178  
 DB 119 LVTND-----NQRVQGFSTDEKGNIVNKMTDIKPPR-ALIPAKATKELKLDNLDSRM 170  
 QY 179 DKTQSEANPYFALLSSWKNGNTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSTG 238  
 DB 171 EPTK-----KFPDADPYSTSHYSTG--VEMYDSQGNKHLVSFFP-----NKVN 211  
 QY 239 SKTFEY--LVAMNPSDEGSAAGTDSAGL--LMSGTMTFSSNGELKNM-----TAFPTGS 290  
 DB 212 DREWEFKGLV-----DGKEITGEGEKMSVAAGKLMFTVDGKLDSEQTSTNPNFKGG 265  
 QY 291 ATKDLNMQAPLVLNGLPQFSANFVGAGIQLPTLDFG--IKSQQNMWAGAPASAAAI 348  
 DB 266 ALQD-----QQVKNLFGDAIK-----D 282  
 QY 349 IGKLPSPMPQIOTSSGNSTARNSSSTRYS-----QDGYPGDLVDVTTITSEGLK 399  
 DB 283 GKK-----GLDGTQYKGNKSLISWHQDGAAGTITGLSFNDEGTLT 324  
 QY 400 GKYSNQVVDVFNPIPLAFTSEDLRREGNNHYSATLDSGGPEPGLPSTNYGKLSVNQL 459  
 DB 325 AVYSNGQANDLAQIALAKFENPEALFKVGNRLKESRDSGTASVAGAPGAAGRGKLF 384  
 QY 460 ETSNVDMREVMNMIIOQGFQMSKSVTTADTLMQKALELKR 502  
 DB 385 ERSTVDLATEFVNMIQORGFQANAKITTTTDELLNEVIQLR 427

## RESULT 11

FLGE TREPH STANDARD; PRT; 463 AA.  
 ID FLGE TREPH STANDARD; PRT; 463 AA.  
 AC Q56326.  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Flagellar hook protein flgE.  
 GN Name=flgE;  
 OS *Treponema phagedenis*.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; *Treponema*.  
 OX NCBI\_TaxID=162;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Kazan 5;  
 RX MEDLINE=94266716; PubMed=8206841;  
 RA Linberger R.J., Slivinski L.L., Samsonoff W.A.;  
 RT "Genetic and biochemical analysis of the flagellar hook of *Treponema*  
 RT phagedenis".  
 RL J. Bacteriol. 176:3631-3637(1994).  
 CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; U04619; AAA73467.1; --  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
KW Flagellum.  
SQ SEQUENCE 463 AA; 49527 MW; 960548C364CA35BC CRC64;  
  
Query Match 21.9%; Score 567.5; DB 1; Length 463;  
Best Local Similarity 31.2%; Pred. No. 2e-26;  
Matches 162; Conservative 66; Mismatches 217; Indels 75; Gaps 11;  
  
Qy 1 MMSLFGATGKTHSTGLGTVSNINIANITGYKQQVVFODLFSQDLAIGS-----TG 55  
Db 1 MMSLFGVSGVQHQTRMDVIGNVNVNVTGFRGRVNPQDLISQOLSGASRPNVEVG 60  
  
Qy 56 SQGNQAGAGVGSVRTIFQGAPEPGNSVTDLAIGKGFQVTVLEDKVHYTRAGNFRF 115  
Db 61 GVNPEVLGVNVAIDTVHTQALQTTGINTDAIQNGFFILKDGKESFYTTAGAGV 120  
  
Qy 116 TODGFLNDPSGFTLMGSRISNNPIKKTLEPIQI-----DFNDPTVAKSPAKTSTA 167  
Db 121 DRDGLTVNPAN-----GMEVQ---GMAEDIEGQIINTSDQTEDLIPIQKIDAKATTD 173  
  
Qy 168 LNAVNVNGDSTDKTQSEANPYFALLESKWKGTPPISTSNYSYAPMRVYDQOQNSHDI 227  
Db 174 VAYACNLKRLPELPEGANQADILSTW-----ATDENVYDTFGQEHKLQ 218  
  
Qy 228 VYFDGAPSGSTKTFEYLVANPNPEEDGSAAGTDSAGLLMSGT-----MTFSSNGELKNM 282  
Db 219 NVFSRVPCTNN-----QWLATVNDPENQAGTETRVGIGTTDGTENTFTVSPDNYCHLASV 274  
  
Qy 283 TAPFTGATKDLNMQAPLVNGLPQPSANFVNGAGIPLDLFCIKSQQNMWAGAPASA 342  
Db 275 T--DTAGNVT-----APAGQVLVQASYNVVGA-----NPDEGGAPTR 309  
  
Qy 343 AAGITDGLKPSMMPITQSSNSTARNSSSTRYSODGYPQGDLDVDTITSEKLOCKY 402  
Db 310 HTFENILGE-----IGTSRNTITQFAERSTTKAYQDGYAMGYLENFKIDQSGIITGVY 363  
  
Qy 403 SNSQVDFNIPDAPFTSGEDGLRRGNHYSATLDSGGPEFGLGTSNGKLSVNOLETS 462  
Db 364 SNGASREIQALAGFANQGGLEKAGENTYIQSNNSGANITVSGVMGKGLIAGTLEMS 423  
  
Qy 463 NYVDSRMVNMIIQRFQMSKSVTTADTLMOKALELKR 502  
Db 424 NVDLTDQFTDMITQRGFOAGAKTIQTSDTMTLETVLNLR 463  
  
RESULT 12  
Q8ZFB5 PRELIMINARY; PRT; 428 AA.  
AC Q8ZFB5; Q74UV4; Q7CHX9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DE Flagellar hook protein FlgE (Hook protein).  
GN Name=flgE; Synonyms=Fla FV, flak, flgE1;  
GN OrderedLocusNames=YP1591, YPO1802, Y2507;  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;  
RA Parkhill J., Wren B.W., Thomson N.R., Tittball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Farraga A.-M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Fellwell T., Hamlin N., Hollroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
RL Nature 413:523-527(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIMS / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM.";  
RL J. Bacteriol. 184:4601-4611(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=91001 / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,  
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,  
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,  
RA Yang R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ414150; CAC90619.1; --  
DR EMBL; AB013953; AAM80604.1; --  
DR EMBL; AB017132; AAS61825.1; --  
DR PIR; AG0219; AG0219.  
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; Flae.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; FlagE; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
KW Complete proteome; Flagellum.  
SQ SEQUENCE 428 AA; 44875 MW; 2FC6C6CF0A255055 CRC64;  
  
Query Match 21.6%; Score 557.5; DB 2; Length 428;  
Best Local Similarity 29.1%; Pred. No. 7.2e-26;  
Matches 145; Conservative 74; Mismatches 196; Indels 83; Gaps 11;  
  
Qy 10 TGMKTHSTGLGTVSNINIANITGYKQQVVFODLFSQDLAIGTSGQPNQAGMAYG 69  
Db 8 SGMAASSNLDVIGNNINIANISATSGFKAGSVFADMPA-----GS-----QTGMGVKVA 55  
  
Qy 70 SVRTITQGAPEPGNSVTDLAIGKGFQVTVLEDK-----VHYTRAGNFRFTQDGLNDPSG 126  
Db 56 GITQDFNDGTATTNRLDLAISQNGFFR--MQDSSGGIYAAENGQFKLDENENIVNMQG 113  
  
Qy 127 FTLMGSRISNNPIKKTLEPIQIOLFNDPTVAKSPAKTALNAVNLGDSSTDKTQSEAN 186  
Db 114 LNLTYPATCTPTVQGANFVPLSIPQDMIS---AKATTSGNMVANLTSTHVDIAEATS 170  
  
Qy 187 PYFALLESWKGNGTTPPISTSNYSYAPMRVYDQOQNSHDIIV-YFDGAPSGSTGKTFEYL 245  
Db 171 PF-----DPPNPPTYSFVNNMTTFDSLGRHEINLVYVRAEDATDGTNDVY 218  
  
Qy 246 VAMNPSBDGSAAGTDSAGLLNSGTMTTFSSNGELKKNMTAFTPTGSAATKDLNMQAPLVN 305  
Db 219 TRDSSAKVTFDPAQTPDPAARAKRGMVFDNSGALKXVNTGT-NATSTTDTFTTIPMGVN 277  
  
Qy 306 GLP--QFSANFVAGIQTFLDFGIKSGQNMWAGAPASAAIAGTDICKLFSMMPITQSSG 363

```
Db 278 GAPAQSFALNVAGS-----KQON----- 295
Qy 364 NSTARNSSSTRYSQDYPQGGDLVDVITITSEKLGKQKYSNSQVVDYFNIPILARETSEDG 423
Db 296 -----TGADSIVAQNTGTAAAGBEFTGFQINSQDGVVGTYSNQQTLLGQIVMVNFSPNEG 350
Qy 424 LRREGNNHYSATLDSGGPFGFLPCITSNYKLSVNOLETNSVDMRSRMVNMIIQRFQOMN 483
Db 351 LSSEGDVNWKETOSSGNFTLTAGSGGFTLTSGALESSNVDLSKELVNMVIAQRNYQSN 410
Qy 484 SKSVTTADTMLOKALELK 501
Db 411 AQTIKTQDQILQTLVSLR 428

RESULT 13
AAS61825
ID AAS61825 PRELIMINARY; PRT; 428 AA.
AC AAS61825;
DT 24-MAR-2004 (TrEMBLrel. 27, Created)
DT 24-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN FLG1 OR YP1591.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017132; AAS61825.1; -.
KW Flagella.
SQ SEQUENCE 428 AA; 44875 MW; 2FC6CACF0A255055 CRC64;

Query Match 21.1%; Score 557.5; DB 2; Length 428;
Best Local Similarity 29.1%; Pred. No. 7.2e-26;
Matches 145; Conservative 74; Mismatches 196; Indels 83; Gaps 11;

Qy 10 TGMKTHSTGLGTVSNNIANTIGYKQOQVVFQDLFSQDLAIGSTGSGQPNQAGQAVG 69
Db 8 SGMAAASNLDIVGNNIANSATSGPKAGSVSFADMF-----GS-----QTGMGVKVA 55
Qy 70 SVRTFTQGAEPGNSVTDLAIGKGFQVLTLEK---VHYTRAGNFRFTQDGLNPDSPG 126
Db 56 GITQDFNDGATTTNRRLLDLAISONGFPR--WDQSSGGIYVARNQPKLDENRNVNMQG 113
Qy 127 FTLGMSRISNNPNIKKETEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDTQSEAN 186
Db 114 LNLGYPATGPTPTVQOGANFVPLSIPQDMIS---AKATTSGNVANLSTHVDVIAEATS 170
Qy 187 PFALLEKSWKNGTPPISTISNYSAQPNRVVDQOQNSHDITV-YFDGAPSTGSKTPEYL 245
Db 171 PF-----DPDNPDTYSFVNNMTTFDLSGRNHEINVFVYKAEADTGDNTVDVY 218
Qy 246 VAMNPSEGSAASTDSAGLLMGTMTFSSNGELKNMTAFTPTGSATKDLNMAQAPLVN 305
Db 219 TRDSAKVTDADPTDPAARAKRSVDFSDGALKVNTNGT-NAITSTDTFTFIPMGVNV 277
Qy 306 GLP--QFSANFVAGIQPLTDFGKIQNNWAGAPASAAAGTIDGKLPMSMPIQTSSG 363
Db 278 GAPAQSFALNVAGS-----KQON----- 295
Qy 364 NSTARNSSSTRYSQDYPQGGDLVDVITITSEKLGKQKYSNSQVVDYFNIPILARETSEDG 423
Db 296 -----TGADSIVAQNTGTAAAGBEFTGFQINSQDGVVGTYSNQQTLLGQIVMVNFSPNEG 350
```

```
Qy 424 LRREGNNHYSATLDSGGPFGFLPCITSNYKLSVNOLETNSVDMRSRMVNMIIQRFQOMN 483
Db 351 LSSEGDVNWKETOSSGNFTLTAGSGGFTLTSGALESSNVDLSKELVNMVIAQRNYQSN 410
Qy 484 SKSVTTADTMLOKALELK 501
Db 411 AQTIKTQDQILQTLVSLR 428

RESULT 14
Q72T33
ID Q72T33 PRELIMINARY; PRT; 464 AA.
AC Q72T33;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FigE.
GN Name=figE; OrderedLocusNames=LIC11188;
OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar Copenhageni).
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=44275;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Fiocruz L1-130;
RX PubMed=15028702;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A., Carrer H.,
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,
RA Coutinho L.B., Degraive W.M., Dellagostin O.A., El-Dorri H.,
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,
RA Kuramao E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L. Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A.,
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.;
RT "Comparative genomics of two Leptospira interrogans serovars reveals
RT novel insights into physiology and pathogenesis.";
RL J. Bacteriol. 186:2164-2172(2004).
DR EMBL; AB017291; AAS69795.1; -.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flab.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; FlaB; 1.
DR Pfam; PF00460; Flg_bb_rod; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 464 AA; 50349 MW; 7EDD5AA84709F84F CRC64;

Query Match 21.3%; Score 550; DB 2; Length 464;
Best Local Similarity 27.9%; Pred. No. 2.3e-25;
Matches 149; Conservative 84; Mismatches 199; Indels 102; Gaps 12;

Qy 1 NMGSIFIGATGKTHSTGLGTVSNNIANTIGYKQOQVVFQDLFSQDLAIGS-----TG 55
Db 1 MRRSYISGVSGLNKHQVRMDVIGNNISNVNTHGFKTERVTFQDMISQELRGASEPKENIG 60
Qy 56 SQGNQAGMGAGVSVRTFTQGAEPGNSVTDLAIGKGFQVLTLEKDVHYTRAGNFRF 115
Db 61 GYNPQVGLGSLJLAIDKIMTQSLQTTGKNTDVAISGEFFIVKDGKQFYTRAGFNL 120
Qy 116 TQDG-FLNDPSGFTLMG--SRISNNPNIKKETEPIQLDFNDPTVAKSPAKTSTALNAV 172
Db 121 DKNGYVNPANGLKVGWNSRLDDKGNKYINSAASIE-DIIPVYSKEPARATSIDPKS 179
Qy 173 NLGSDT-----DKTQSEANPYFALLEKSWKNGTPPISTISNYSAQPNRVVDQOQNSHDIT 227
Db 180 NLNSAPAVFPDATQEEIT--AMIND-----PDKMRGRHVTITNTFDQGIQREFK 229
Qy 228 VTFDGAPESTGSKTPEYLVAMNPSEDSGAAGTDSA-----GLMSTGVTFFSS--- 275
```

Db 230 MEFYKVDNTWKARLSMTDSTQLSDVSGTGQNTQLPGNTELEFGFTPDGKLVVSDGV 289  
Qy 276 -----NGELKNMTAFTPTGS---ATKDLNAPAPLVNGLPQFSANFVGAGIQTLTLDLFGI 328  
Db 290 DSMNSGKLNARVSPFIPGNPAIONFDNLGE-AGVNGITQFSSDF----- 334  
Qy 329 KSQONWAGAPASAAAIGTDIGKLPSPWPIOTSGNSTARNSSSTRYSQDGPQGLV 388  
Db 335 -----TTKAVKQDGYTNGYLE 350  
Qy 389 DVITITSEGLQKYSNQVDFYNIPLARTSEDGLRREGNNHYSATLDSGGPEFGLPQT 448  
Db 351 SFSIDNSGTTGVFSGVROPLARVATAVFNPAGLDKAGDMFYSYNSGEPNIGEAGV 410  
Qy 449 SNYKLSVNOLETSDVMSREVMNMIIOQGFQMNKSVTTADTMQKALELKR 502  
Db 411 QGRGKINAGLLENSVDLSQFTDMIVTQRFQANSRTITTSQMIQEVGLKLR 464

RESULT 15  
Q8F2C6  
ID Q8F2C6 PRELIMINARY; PRT; 464 AA.  
AC Q8F2C6  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)  
DE Flagellar hook protein.  
GN Name=flag; OrderedLocusNames=LA2848;  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroaceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN NCBI\_TaxID=173;  
RP SEQUENCE FROM N.A.  
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
RX MEDLINE=22598143; PubMed=127112204;  
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,  
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-P., Jiang H.-O., Jia J., Tu Y.-F.,  
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,  
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-X., Ma W.,  
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,  
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,  
RA Xu J.-G., Zhao G.-P.;  
RT "Unique physiological and pathogenic features of Leptospira  
RT interrogans revealed by whole-genome sequencing."  
RL Nature 422:888-893(2003).  
DR EMBL; AE011450; AAN50047.1;  
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; P:motor activity; IEA.  
DR GO; GO:0005198; P:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; Flae.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; Flae; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00368; FLAGELLA\_BB\_ROD; UNKNOWN\_1.  
KW Complete proteome; Flagellum.  
SQ SEQUENCE 464 AA; 50349 MW; 7EDD5AA84709F84F CRC64;

Query Match 21.3%; Score 550; DB 2; Length 464;  
Best Local Similarity 27.9%; Pred. No. 2.3e-25;  
Matches 149; Conservative 84; Mismatches 199; Indels 102; Gaps 12;

Qy 1 MMSGLFTGATGKTHSTGLGTSSNNANANTIGYKQOVVFDLFSQDLAIGS-----TG 55  
Db 1 MMSRLSYGVSGGLKNNHQRMDVIGNNSVNTHTGFKTERVTFQDMISQELRGAEPKENIG 60  
Qy 56 SQGNQAGMGAQVGSVRTITFQGAPEPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRF 115  
Db 61 GVNPOQVGLSLIAAIDKIMTQSLQTTGKNTDVAISGEGFIVKDGKQFYTRAGFNL 120  
Qy 116 TDQD-FLNDPSGFTLMG--SRISNNPNIKKTELEIQLDFNDPTVAKSPAKTSTALNAV 172

Db 121 DKNGYVNPANGKLVQGWNSRLDDKGNKYINSAASIE-DIIIPVYSKEPARATSOIDFKS 179  
Qy 173 NLGDS-----DKTQSEANPYFALLEWSKMGNGTPISTSNYSYAQPMVYDQGNSHDIT 227  
Db 180 NLSAPAPVPPDATOEIT---AMIND-----PDKWRGHVTTINTFDDQGIQREFK 229  
Qy 228 VYFDGAPSTGSKTPEYLVAMNPSDEGSAASGTSA-----GLMSGTMFSS---- 275  
Db 230 MEFYKVDNTWKARLSMTDSTQLSDVSGTGQNTQLPGNTELEFGFTPDGKLVVSDGV 289  
Qy 276 -----NGELKNMTAFTPTGS---ATKDLNAPAPLVNGLPQFSANFVGAGIQTLTLDLFGI 328  
Db 290 DSMNSGKLNARVSPFIPGNPAIONFDNLGE-AGVNGITQFSSDF----- 334  
Qy 329 KSQONWAGAPASAAAIGTDIGKLPSPWPIOTSGNSTARNSSSTRYSQDGPQGLV 388  
Db 335 -----TTKAVKQDGYTNGYLE 350  
Qy 389 DVITITSEGLQKYSNQVDFYNIPLARTSEDGLRREGNNHYSATLDSGGPEFGLPQT 448  
Db 351 SFSIDNSGTTGVFSGVROPLARVATAVFNPAGLDKAGDMFYSYNSGEPNIGEAGV 410  
Qy 449 SNYKLSVNOLETSDVMSREVMNMIIOQGFQMNKSVTTADTMQKALELKR 502  
Db 411 QGRGKINAGLLENSVDLSQFTDMIVTQRFQANSRTITTSQMIQEVGLKLR 464

RESULT 16  
AAS69795  
ID AAS69795 PRELIMINARY; PRT; 464 AA.  
AC AAS69795;  
DT 31-MAR-2004 (TremBLrel. 27, Created)  
DT 31-MAR-2004 (TremBLrel. 27, Last sequence update)  
DT 11-MAY-2004 (TremBLrel. 27, Last annotation update)  
DE Flge.  
GN FLGE OR LIC11188.  
OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar  
OS Copenhagen).  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroaceae; Leptospira.  
OX NCBI\_TaxID=44275;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fiocruz L1-130;  
RX PubMed=15028702;  
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,  
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,  
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrier H.,  
RA Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H.,  
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Gigliotti E.A.,  
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,  
RA Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,  
RA Kuranae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,  
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A.,  
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,  
RA Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.;  
RT "Comparative genomics of two Leptospira interrogans serovars reveals  
RT novel insights into physiology and pathogenesis."  
RL J. Bacteriol. 186:2164-2172(2004).  
DR EMBL; AE017291; AAS69795.1;  
SQ SEQUENCE 464 AA; 50349 MW; 7EDD5AA84709F84F CRC64;

Query Match 21.3%; Score 550; DB 2; Length 464;  
Best Local Similarity 27.9%; Pred. No. 2.3e-25;  
Matches 149; Conservative 84; Mismatches 199; Indels 102; Gaps 12;

Qy 1 MMSGLFTGATGKTHSTGLGTSSNNANANTIGYKQOVVFDLFSQDLAIGS-----TG 55  
Db 1 MMSRLSYGVSGGLKNNHQRMDVIGNNSVNTHTGFKTERVTFQDMISQELRGAEPKENIG 60  
Qy 56 SQGNQAGMGAQVGSVRTITFQGAPEPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRF 115  
Db 61 GVNPOQVGLSLIAAIDKIMTQSLQTTGKNTDVAISGEGFIVKDGKQFYTRAGFNL 120

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

Query Match 20.7%; Score 534.5; DB 1; Length 463;  
 Best Local Similarity 28.1%; Pred. No. 2e-24;  
 Matches 154; Conservative 80; Mismatches 183; Indels 131; Gaps 14;

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

Query Match 20.7%; Score 534.5; DB 1; Length 463;  
 Best Local Similarity 28.1%; Pred. No. 2e-24;  
 Matches 154; Conservative 80; Mismatches 183; Indels 131; Gaps 14;

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

CC 116 TDGQ-FLNDPSTGFTLMS--SRISNNPNKIKETLEPIQLDNDFTVAKSPAKTSTALNAV 172  
 DR EMBL; AE001244; AAC65692.1; --  
 DR PIR; C71292; C71292.  
 DR TIGR; TP0727; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb\_rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb\_rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 DR Complete proteome; Flagellum.  
 KW SEQUENCE 463 AA; 49178 MW; 94704C413B36E7A8 CRC64;  
 SQ

|                                                                    |                                                                                                                                    |
|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| GN                                                                 | Name=FlgE3; OrderedLocusNames=TTE1433;                                                                                             |
| OS                                                                 | Thermoanaerobacter tengcongensis.                                                                                                  |
| OC                                                                 | Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;                                                                         |
| OC                                                                 | Thermoanaerobacteriaceae; Thermoanaerobacter.                                                                                      |
| OX                                                                 | NCBI_TaxID=119072;                                                                                                                 |
| RN                                                                 | (1)_TaxID=119072;                                                                                                                  |
| RP                                                                 | SEQUENCE FROM N.A.                                                                                                                 |
| RC                                                                 | STRAIN=MB4.                                                                                                                        |
| RY                                                                 | MEDLINE=21992816; PubMedId=1197336;                                                                                                |
| RA                                                                 | Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,                                                                   |
| RA                                                                 | Chen Y., Xue Y., Xu Y., Bai X., Huang L., Dong X., Ma Y., Ling L.,                                                                 |
| RA                                                                 | Tan H., Chen R., Wang J., Yu J., Yang H.;                                                                                          |
| RT                                                                 | "A complete sequence of the T. tengcongensis genome.";                                                                             |
| RL                                                                 | Genome Res. 12:689-700(2002).                                                                                                      |
| DR                                                                 | EMBL; AE013102; AAM24655.1; ..                                                                                                     |
| DR                                                                 | GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.                                                                                 |
| DR                                                                 | GO; GO:0003774; F:motor activity; IEA.                                                                                             |
| DR                                                                 | GO; GO:0005198; P:structural molecule activity; IEA.                                                                               |
| DR                                                                 | GO; GO:0001539; P:ciliary/flagellar motility; IEA.                                                                                 |
| DR                                                                 | InterPro; IPR010930; DUF1078.                                                                                                      |
| DR                                                                 | InterPro; IPR011491; FlaE.                                                                                                         |
| DR                                                                 | InterPro; IPR001444; Flag_bb_rod.                                                                                                  |
| DR                                                                 | Pfam; PF06429; DUF1078; 1.                                                                                                         |
| DR                                                                 | Pfam; PF07559; FlaE; 1.                                                                                                            |
| DR                                                                 | Pfam; PF00460; Flg_bb rod; 1.                                                                                                      |
| DR                                                                 | PROSITE; PS00589; FLAGELLA_BB_ROD; 1.                                                                                              |
| KW                                                                 | Complete proteome; Flagellum.                                                                                                      |
| SQ                                                                 | SEQUENCE 415 AA; 43973 MW; 9EC16SE38510ECF0 CRC64;                                                                                 |
| <br>                                                               |                                                                                                                                    |
| Query Match            20.4%; Score 528.5; DB 2; Length 415;       |                                                                                                                                    |
| Best local similarity    28.9%; Pred. No. 4e-24;                   |                                                                                                                                    |
| Matches 148; Conservative 77; Mismatches 180; Indels 107; Gaps 12; |                                                                                                                                    |
| QY                                                                 | 1 MMGSLFIGATGMKTHSTGLTGVNNANANTIGYKQQVVFODLSQDL----AIGSTG 55<br>          :           :           :           :           :        |
| DB                                                                 | 1 MLRMYSAISLOAHQARLDVGNIANNVTGYKASMTFKETISQIKGSAPQGNGG 60<br>          :           :           :           :           :           |
| QY                                                                 | 56 SQGPNAQMGAGVSVRIHTFGAFPFGNSYTDLAIGKGKFFQVLTLEDKVHYTRAGNFRP 115<br>          :           :           :           :           :   |
| DB                                                                 | 61 GTNPQQICLGVAIASIDTLFTRGCAORTNPDTLSDGNGGFFIVSGESILYTRAGNESF 120<br>          :           :           :           :           :   |
| QY                                                                 | 116 TDGFLNDPSGFTLMGSRISNNPNKKET--LEPIOLD--FN-DPTVAKSPAKTTALNA 170<br>          :           :           :           :           :   |
| DB                                                                 | 121 DSNGDLVTPGGYKVLGWMSDGTGVNTDGNLVPLSKNWFGWEPS-----TTTQL 173<br>          :           :           :           :           :       |
| QY                                                                 | 171 VNLGDSTDYTOSEANPYFALLESWKXNGTPTPISINYSYAOPMRYVYQQGNSHDIITYF 230<br>          :           :           :           :           : |
| DB                                                                 | 174 GGNNLASTGTGCQS-----ISY--NIIVYTGQGHVAITTF 207<br>          :           :           :           :           :                    |
| QY                                                                 | 231 DGAPSSGTGKTPEYLVMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNWTAFPTGTS 290<br>          :           :           :           :           :  |
| DB                                                                 | 208 TRQDPNTWN-----WSVSSQDPYISSVSAG--STAITFGADGIAPNT--QATGT 254<br>          :           :           :           :           :      |
| QY                                                                 | 291 ATKDLNAWQAPLVNGLPOFSANFVGAGIQPLTDFGKSCQNWMAGAPASAAAITGDIG 350<br>          :           :           :           :           :   |
| DB                                                                 | 255 LTFFNMNT-----AVTNAQIGPVNIDLS 276<br>          :           :           :           :           :                                |
| QY                                                                 | 351 KLPMMFPLOTSSGNSTARNGSSSTRYSODGYPQGDVDVITTSKGLGKYNSQVVDF 410<br>          :           :           :           :           :     |
| DB                                                                 | 277 KL-----TWFTSTDLRELSKNGEAGLSQINDIXGVVSYSGIYNGRQVI 323<br>          :           :           :           :           :            |
| QY                                                                 | 411 YNIPLARFTSDELRRGNHHYSATUDSGPFEGLPGTSNYGKLSNQOLETSNVDMKSEM 470<br>          :           :           :           :           :   |
| DB                                                                 | 324 GQIALADFQPMGLEKVGCTMTINTVNSEPMIGAAGSGTRGSINPGTLEMSVNLANEF 383<br>          :           :           :           :           :   |
| QY                                                                 | 471 VNMIIRGFQMNSKSVTTADTMLQKALEKR 502<br>          :           :           :           :           :                               |
| DB                                                                 | 384 VDMITTQRGFQANAKVITVSDEMLQDLVNMR 415<br>          :           :           :           :           :                             |
| <br>                                                               |                                                                                                                                    |
| RESULT 20                                                          |                                                                                                                                    |
| Q914P9 PRELIMINARY; PRT; 462 AA.                                   |                                                                                                                                    |
| ID                                                                 | Q914P9                                                                                                                             |

Q914P9;  
 01-MAR-2001 (TrEMBLrel. 16, Created)  
 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Flagellar hook protein FlgE.  
 Name=flgE; OrderedLocNames=PA1080;  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Garber R.L., Goltzer L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964 (2000).  
 DR EMBL; AE004539; AAC04469.1; -.  
 DR PIR; F83510; F83510.  
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
 DR GO; GO:0003774; F:motor activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR011491; Flae.  
 DR InterPro; IPR001444; Flag\_bb rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF07559; Flae; 1.  
 DR Pfam; PF00460; Flg\_bb rod; 1.  
 KW Complete proteome; Flagellum.  
 SQ SEQUENCE 462 AA; 48336 MW; 519616DABACE9C48 CRC64;  
 Query Match 20.4%; Score 527.5; DB 2; Length 462;  
 Best Local Similarity 28.5%; Pred. No. 5.4e-24;  
 Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;  
 4 SLFIGATGKHTSTGLGTGVSNNIANITGYKQOVVFQDLFSQDLAIGSTGSGPQAG 63  
 2 SFNIGLSGQAASSGLANTGNIANAGTVGKQRAEFADYAAASV-LGS-----GSPQ 56  
 64 MGAQVGSVRTIPTQGAFFPGKSVTDLAIGKGFQVLTEDKVHYTRAGNPRFTODGFLND 123  
 57 SGVLLSDVSQMPKQGNIDSTNSVLDAINGNGFFVTSNNGAISTRAGYFNTDKQFIVD 116  
 124 PSGFILMSRISNNPIKKEFILEPQLD-FNDPTVAKSPAKTSPALNAVNLGSDTDKQ 182  
 117 NNGYRLOGYAVGPNQQLQGVVTDLKVQRANQAPQATSSIQSYNLNSTLK-----P 168  
 183 SEANPYFALLSWKNGPPTSTNYSVAQPMRVVDQGNSHDITVFE---DGAPEST-- 237  
 169 PTVTFP-----DPSDATYNSSSGLGYDSGNHSTMSQFFIKNEPDPNATPP 216  
 238 ---GSKTFEYLV-AWNPESDGAASGSDSAGLLSGMTWTFSSNGELKNMTAFTPTGS--- 290  
 217 IPENSWTKVLIDGVNPLDPSNKT-----MSFNVTDPASQMTSVRA---PDGSTG 266  
 291 -----ATKDLNAPPLVNLGPFQSANFVGAGIQPLTLDLFGKQSNWAGAPASAAA 344  
 267 PGFIDATNVIQSPA---TCNPPTP-----GTGWIPAAASD-GKTPPTPYAWNGATGAAG 318  
 345 IGTDIGKLPSMMP:QTSGSNSTARNGSSSTRYS-----QDGYPGQDLVDVITSE 395  
 319 ISFVWRK-----TTQYSTAFQAQSNPIQDGYTTCQLAGLEIDDT 356  
 396 GKLOGKYSNOVDYFNIPILPFTSEDLRREGNNHYSATLDSGGPFGPLGTSNTGKLS 455  
 357 GVIFARYTNGSKVQGGVLANFANIQGLTPIGKTSWVQSSGESGPAVGAPRSGTIGALQ 416

QY 456 VQOLETNSVDMRSREVMNMIIRQGFOMNSKSVTTADTMLOKALELK 501  
 DB 417 SGALEASNVDSINELVNLVHQRYNQANAKTIQTEDAVTQTIINLR 462  
 RESULT 21  
 FLGE\_BORBU  
 ID FLGE\_BORBU STANDARD; PRT; 442 AA.  
 AC Q44767; Q44734; Q44899;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Flagellar hook protein flgE.  
 GN Name=flgE; OrderedLocNames=BB0283;  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RA Dunn J.J., Butler-Loffredo L., Kieleczawa J., Medalle J., Luft B.J.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HB19;  
 RA Old I.G.;  
 RN [3]  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,  
 Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,  
 Richardson D.L., Peterson J.D., Keflavage A.R., Quackenbush J.,  
 Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,  
 Gocayne J.D., Weidman J.F., Utterback T.R., Wattley L., McDonald L.A.,  
 Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,  
 Roberts K.M., Hatch B., Smith H.O., Venter J.C.;  
 "Genomic sequence of a Lyme disease spirochaete, Borrelia  
 burgdorferi.";  
 RT Nature 390:580-586 (1997).  
 [4]  
 RP SEQUENCE OF 18-207 FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RA Limberger R.J., Slivenski L.L.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins  
 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; U43739; AAA85606.1; -.  
 DR EMBL; U43849; AAA87351.1; -.  
 DR EMBL; AE001137; AAC66665.1; -.  
 DR EMBL; U19712; AAA61738.1; -.  
 DR PIR; C70135; C70135.  
 DR TIGR; BB0283; -.  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR001444; Flag\_bb rod.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF00460; Flg\_bb rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 KW Complete proteome; Flagellum.  
 FT VARIANT 18 18 R -> I (in strain HB19).  
 FT VARIANT 119 119 D -> Y (in strain HB19).

FT VARIANT 174 174 V -> I (in strain HB19).  
FT VARIANT 192 192 N -> S (in strain HB19).  
SQ SEQUENCE 442 AA; 47389 MW; 561AC092B72C69BE CRC64;  
  
Query Match 19.7%; Score 509; DB 1; Length 442;  
Best Local Similarity 28.2%; Pred. No. 6.7e-23;  
Matches 150; Conservative 71; Mismatches 190; Indels 120; Gaps 13;  
  
QY 1 MMSLFIGATGKMTHTGLTGVSNINANTIGYKQQVFDLPFSDLAIGS-----TG 55  
Db 1 MMSLSYGVSGLQNHQTHQMDVGVGNINANTIGYKQVFDLPFSDLAIGS-----TG 55  
QY 56 SQCPNQAAGCAQVGVRTFTTQGAPEPGNSVTDLAIGGKGFQVLTEDKHYTRAGNPRF 115  
Db 61 GTNPKQVGLGMNVASIDTHTTQGAQSTOKASDLGVSGNGFFLKKEGKNLYTRAGAFDV 120  
QY 116 TQDGLNDPSGFTLMGSRISN--NPNIKKETLEPIQLDND---PTVAKSPAKTSTALNA 170  
Db 121 DSDRLHVNPN-----GMRIQGWARDLEGEKIVNTASDIEDLIIPIGKEGAKTKNVT 176  
QY 171 VNLGSDTDKTOSEANPYFALLESWKNGTTPRISTSNYSYAQPMRVYDQGNSHDITVYF 230  
Db 177 ACNLDKRLPLIQEGANPADIRGTWVVKNS-----LYDSFGN---VSVL- 217  
QY 231 DGAPSTGSKTEFYLVAWNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNWTAFTPTGS 290  
Db 218 -----ELR----- 220  
QY 291 ATKDLNA--WOPAPLVNGLPQFSANFVGAGTQPLTLPF-----GIKSOQNMWAG 337  
Db 221 VVKDLATPNLWATVLNG--EQNSNF-----TLGPDNEGALASLNGPGQKGDILQ 270  
QY 338 APASAAAIGTDIGKLSMMPQIOTSSG-----NSTARNGSSSTRYSDQGYPOGDLVDV 390  
Db 271 IPITFNVLGANVEGEQQTQVNLKLTGVSTYDTSITQFADSSSTKAIITQDGYGMGYENY 330  
QY 391 TITSEGLQKYSNSQVDFYNIPLARFTSEGLRREGNNHYSATLDSGGPEFGFLPGTNS 450  
Db 331 EIDQNGVIIVGYSNGIRRLGXIALASFWNPGLAKSGDTNFVETNSGQVRIGETGLAG 390  
QY 451 YGKLSVNOLETNSVDMRSREWNMIIOQGFQWNSKSVTTADTMLOKALELK 501  
Db 391 LGSIRSGVLEMANVDLBOFTDMIVTQRFQANAKTITTSOLLQELVRLK 441  
  
RESULT 22  
Q8XSX2 PRELIMINARY; PRT; 401 AA.  
AC Q8XSX2  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE PROBABLE FLAGELLAR HOOK PROTEIN FLGE.  
GN Name=flgE; Synonyms=RS00745; OrderedLocusNames=RS0345;  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OG Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GMI1000;  
RX MEDLINE=21681879; PubMed=11823952;  
RA Saranoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Ariat M., Billaut A., Brottier P., Camus J.C., Catolico L.,  
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
BL Nature 415:497-502(2002).  
DR EMBL; AL646078; CAB17496.1;  
GO; GO:0009288; C:Flagellum (sensu Bacteria); IEA.

DR GO; GO:0003774; E:motor activity; IEA.  
DR GO; GO:0005198; P:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; FlaeE.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; FlaeE; 1.  
DR Pfam; PF00460; Fig\_bb\_rod; 1.  
DR PROSITE; PS00888; FLAGELLA\_BB\_ROD; 1.  
KW Complete proteome; Flagellum; Plasmid.  
SQ SEQUENCE 401 AA; 41576 MW; 6F18569259DACC0 CRC64;  
  
Query Match 19.4%; Score 502; DB 2; Length 401;  
Best Local Similarity 27.4%; Pred. No. 1.6e-22;  
Matches 137; Conservative 85; Mismatches 168; Indels 110; Gaps 15;  
  
QY 8 GATGKHTHTGLTGVSNINANTIGYKQQVFDLPFSDLAIGSQTSGQPNQAGCAQ 67  
Db 6 GLSGDLAASKNLDVIGSNVANANTVGYKSTAEFGDYARSL-VGAS-----DNQIGQVD 60  
QY 68 VGSVRTFTQGAPEPGNSVTDLAIGGKGFQV--TLEDKHYTRAGNFRFTQDGLNDPS 125  
Db 61 VTKVSOSTQGNVTVTGNPLDIAINGTGYRMDVASSQVSYTENGQFQDKNGYIISAT 120  
QY 126 GTFLMGSRISNPNKIKETLEPIQLDNDPTVAKSP-AKTSTALNAVNVNLDGSDTKTOSE 184  
Db 121 GQNLGTYGVDAATKINTAVLTNLQIPVND-----LAPLATNTAFS---INL-DAAGTVPT- 172  
QY 185 ANPYFALLESWKNGTTPRISTSN--YSYAQPMRVYDQGNSHDITVYFDGAPSTSGTK 241  
Db 173 -----TTFPSATNSATFNHVSVEQYDGTGTSMLTNYV--VRTAAGWDV 215  
QY 242 FRYLVAMPSEDSASGTDAGLLMSGTMTFSSNGELKNWTAFTPGSATKDLNAPQA 301  
Db 216 YSQVDPGANPTGPNVTS-----LTFNSSGVL-----TSSPSKVAFA- 252  
QY 302 PLVNLGLPQFSANFVGAGTQPLTLPFGIKSQQNMWAGAPASAAICTDIGKLSMMPQIOTS 361  
Db 253 -----AGNSIANMDF-----TGT----- 265  
QY 362 SGNSTARNGSSSTRYSDQGYPOGDLVDVTTITSEGLQKYSNSQVDFYNIPLARTSE 421  
Db 266 -----TQYGGGFNDTTSQDGYATGRLASYSVGTDTTIGTYSNGRTSTLGLQIANTFKAP 321  
QY 422 DGLRREGNNHYSATLDSGGPEFGFLPGTNSYKLSVNOLETNSVDMRSREWNMIIOQGFQ 481  
Db 322 DGLQNLGNGQWVETAEASGAPQWGPNGSGFGLQSSAVEQSNVDLSAELVNMIVAQRSYQ 381  
QY 482 MNSKSVTTADTMLOKALELK 501  
Db 382 ANAQTKTEDTILQTLVSMR 401  
  
RESULT 23  
Q8FTQ4 PRELIMINARY; PRT; 401 AA.  
AC Q8FTQ4  
DT 01-MAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Flagellar hook protein flgE.  
GN Name=flgE; OrderedLocusNames=cl345;  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasco D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,



Query Match 19.3%; Score 498; DB 1; Length 402;  
 Best Local Similarity 28.7%; Pred. No. 2.8e-22;  
 Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;

QY 10 TGMKTHSTGLTVSNINANTIGYKQOVVQDLFSQDLAIGSTGSGQPNQAGVQV 69  
 DB 7 SGLNAAATNLDVIGNNIANSATYGFKSGTASADNFA-----GS-----KVLGKVA 54

QY 70 SVRTITQGAFFBEGSVTDLAIGGKGFQ-VLEDKVHYTRAGNFRFTQDGLNDPSGFT 128  
 DB 55 GITQDFTDGTTCRGDLVAISQNGFFRLVDSNGSVFYSRNGQPKLDENRNLVNMQMG 114

QY 129 LMGSRISNNPNKKTLEPIQDLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPY 188  
 DB 115 LTGYPATGTPPTIOQGANPAPITPNTLMA---AKSTTTASMQINL-NSTDVPSKT--- 167

QY 189 FALLESWKGNGTPPISTSTSYVAQPMRVYDOQGNSHDITVYFDGAPSGTSGKTFEYL 245  
 DB 168 -----PFSVDADSYNKKGVTVYVDSQGNADNMVYF-----VKTQDNEWA 208

QY 246 VAMNPSEDSAGSAGTDSAGLLSGTMTFSSNGELK-----NMTAFTPTGSAKDNLNAQP 300  
 DB 209 VYTHDSDDPAAPTTA-----STLKFENGLISGGTVNITTTGTINGATA----- 255

QY 301 APLVNLGPOFSANFVGAGIQPLTDFGIKSOQNMWAGAPASAAAIGTDIGKLPSPMPQT 360  
 DB 256 -----ATFSLSFLNS-----MQQN-----T 270

QY 361 SSGNSTARNSSSTRYSODGYPOGDLVDVITTSSEKLGKYSNSQVVDVFNIPARFTS 420  
 DB 271 GANNIVATN-----QNGYKPGDLVSQYQINNDGTVVGYSNEQEOVLQIVLANFAN 321

QY 421 EPOLRREGNHYSATLDSGGPFGLPGTSNYGKLSVNOLETSNVDMSEVMNMIIOQGF 480  
 DB 322 NEGLASQGNVWAATQASGVALLGAGTSGNFGKLTNGALEASNDLSKELVNMIVAQENY 381

QY 481 QMNSKSVTTADTMLOKALELK 501  
 DB 382 QSNAGTIKTQDQILNTLVNLR 402

## RESULT 25

Q7AF93 PRELIMINARY; PRT; 401 AA.

ID Q7AF93  
 AC Q7AF93;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE Flagellar hook protein flgE.  
 GN OrderedLocustNames=EC81454;  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 CX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Tada T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AP002555; BAB34877.1; --  
 DR InterPro; IPR010930; DUF1078.  
 DR InterPro; IPR011491; Flae.  
 DR Pfam; PF06429; DUF1078; 1.  
 DR Pfam; PF07559; Flae; 1.  
 DR Pfam; PF00460; Flg\_bb rod; 1.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.

KW Flagellum.  
 SQ SEQUENCE 401 AA; 42014 MW; 453B39179717CB14 CRC64;

Query Match 19.1%; Score 493; DB 2; Length 401;  
 Best Local Similarity 28.7%; Pred. No. 5.6e-22;  
 Matches 142; Conservative 76; Mismatches 174; Indels 102; Gaps 15;

QY 10 TGMKTHSTGLTVSNINANTIGYKQOVVQDLFSQDLAIGSTGSGQPNQAGVQV 69  
 DB 8 SGLNAAATNLDVIGNNIANSATYGFKSGTASADNFA-----GS-----KVLGKVA 55

QY 70 SVRTITQGAFFBEGSVTDLAIGGKGFQ-VLEDKVHYTRAGNFRFTQDGLNDPSGFT 128  
 DB 56 GITQDFTDGTTCRGDLVAISQNGFFRLVDSNGSVFYSRNGQPKLDENRNLVNMQGLQ 115

QY 129 LMGSRISNNPNKKTLEPIQDLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPY 188  
 DB 116 LTGYPATGTPPTIOQGANPAPITPNTLMA---AKSTTTASMQINL-NSSDPLPSV--- 167

QY 189 FALLESWKGNGTPPISTSTSYVAQPMRVYDOQGNSHDITVYFDGAPSGTSGKTFEYL 248  
 DB 168 -----NAFDASNADSYNKKGVTVYVDSQGNADNMVYF-----VKTGDNHNVYV--- 211

QY 249 NPSEDSAGSAGTDSAGLLSGTMTFSSNGEL-KMTAFTPTGSAKDNLNAQVAPLVNGL 307  
 DB 212 --TQSDSDPTGTABPAM-----KLVENANGVLTSNPTENITGA-----INGAEP--- 255

QY 308 POFSAFVGAGIQPLTDFGIKSOQNMWAGAPASAAAIGTDIGKLPSPMPQTSSGNSTA 367  
 DB 256 -TFSLSFLNS-----MQQN-----TGANNIVA 276

QY 368 RNSGSTRYSODGYPOGDLVDVITTSSEKLGKYSNSQVVDVFNIPARFTSBDGLRRE 427  
 DB 277 T-----TQNGYKPGDLVSQYQINNDGTVVGYSNEQTLQGLQIVLANFANNEGLASE 327

QY 428 GNNHYSATLDSGGPFGLPGTSNYGKLSVNOLETSNVDMSEVMNMIIOQGFQVMSKSV 487  
 DB 328 GDNVWSATQSSGVALLGAGTSGNFGKLTNGALEASNDLSKELVNMIVAQENYQSNAGTI 387

QY 488 TTADTMLOKALELK 501  
 DB 388 KTQDQILNTLVNLR 401

## RESULT 26

Q8X8L1 PRELIMINARY; PRT; 401 AA.

ID Q8X8L1  
 AC Q8X8L1;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 26, Last sequence update)  
 DE Flagellar biosynthesis, hook protein.  
 GN Name=flgE; OrderedLocustNames=z1714;  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 CX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;  
 RX MEDLINE=21074935; PubMed=11208551;  
 RA Ferna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:528-533(2001).  
 DR EMBL; AE005317; AAGS5822.1; --  
 DR PIR; B85670; B85670.  
 DR PIR; F90810; F90810.  
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.



```
275 VAT-----TQGYKPGDLVSYQINDDGTGVVGNYSNQTOLLGQVLNANFANNEGLA 325
426 REGNNHYSAILDSCGPFGLPRTSNYKLSVNOLETNSVDMREVMNMIIOGFQWNSK 485
326 SEGDNVWSATQSSGVALGTAGTGNFTLTNGALEASNDLSKELVNMIQAORNYOSNAQ 385
486 SVTTADTMLOKALEK 501
386 TIKTQDQILNLTNLR 401

RESULT 28
FLGE_HELPJ
ID_FLGE_HELPJ STANDARD; PRT; 718 AA.
AC
O92KYO:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Flagellar hook protein flgB.
GN Name=flgB; OrderedLocusNames=JHP0804;
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
[1]
SEQUENCE FROM N.A.
MEDLINE=99120557; PubMed=9923682; DOI=10.1038/16495;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.B., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RA "Genomic sequence comparison of two unrelated isolates of the human
RA gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins
CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AE001509; AAD06365.1; -
PIR; G71888; G71888.
InterPro; IPR010930; DUF1078.
InterPro; IPR010810; Flagellin.IN.
InterPro; IPR001444; Flag_bb_rod.
Pfam; PF06429; DUF1078; 1.
Pfam; PF07196; Flagellin.IN; 1.
Pfam; PF00460; Flag_bb_rod; 1.
PROSITE; PS00588; FLAGELLA_BB_ROD; FALSE_NEG.
Complete proteome; Flagellum.
SEQUENCE 718 AA; 76280 MW; DFGA0500B8EF481D CRC64;

Query Match 18.8%; Score 485; DB 1; Length 718;
Best Local Similarity 25.0%; Pred. No. 3.8e-21;
Matches 181; Conservative 94; Mismatches 22; Indels 226; Gaps 22;

QY 1 VMGSLFIGATGKHTSTGLGTYSNNIANANTIGYKQOVWFQDLFSQDLAIGS-----TGS 56
DB 1 MURLSWGNGVNGAHOALDIENNNIANVNTGFKYSASFVDMLSQVKLATAPYKNGL 60
QY 57 QGPN--QAGMGAVGSVRTITFGAFEPGNSVTDLAIGKGFQFVTLBKV--HYTRAGN 112
DB 61 AQGNDFSVGLGVGDATTKIFSGQNIQNTDVKTDLAIQDGGFFIISPRGRTNPTRDGE 120
QY 113 FPFQDGFNDPSGFTLM-----GSRISNNNIKKTLEPIQLDFNDPTVAKSPAK 163
DB 121 FLFDQSGLVTTGGLVWQGVNRSGDTGNKGSDDTDLKVDNTGPLENTRIDPGV-NPAR 179
```

```
QY 164 TST--ALNAVNLGDSKTOSTEANPYFALLESWK--GNGTPPISTSNYSYAQPMR----- 215
DB 180 ASNRISMEANLNGRHADQTAA---VFALDSSAKTPSDGINPVYDSGTNLQAQVAEDMGS 235
QY 216 VTDQGN-----SHDITVYFDGA-----PS-----STGSKTTEY 244
DB 236 LYNEGDGALLNENQGIWYSYKSAKXVXKIDLPSEAENSTLENGVKISFTNDSAVRSSTSL 295
QY 245 LVAMNPSEDGSAAG---TDSAGLLMSGTMTFFSNGBLKN----- 281
DB 296 VAAKXAINAVKSGTGLEIAYLDQKQLRENTNELDQEKIKNIIVTQAGTGAFANFLDGDGK 355
QY 282 -MTAF-----TPTGSATKDLNMQPAPLVNGLPQPSANFV-----GAGIQPLTL 324
DB 356 DVTAFKYSYTHSISPNADIGQPTTDEALITQHDANIVKDPSPDLADNYQDSAAISGVTTIN 415
QY 325 DFG-----IKSQNNW-----AGAPASAA 343
DB 416 QYGMFEINKONKNIKENLNIFVSGYSDSVTNVLPKNAMKGLNTASLISSGASASS 475
QY 344 -----AIGT----- 347
DB 476 KETHATHATSIDVIDSLGTKHAMRIEFYRSGGAENFRVIVPEPGELVGGSAARPNVEG 535
QY 348 -----DIGKLPSMP-----IOTSSGNSTARGSSSTRYS-----O 379
DB 536 GRHFNNDSGLAGMNPPLLPQDPKNGADAPQRIINLAFSGSGFDGLTSVDKISFYAIEQ 595
QY 380 DGYVQGDLDVVDVITSEGLQKYSNSQVVDVFNPIPLARTSDEGLRRRGNHYSATLDG 439
DB 596 NGYQAGDLDVDFSDGVLGAFSGRTLALAAQVALANFANDAGLQALGNGVFSQTNSG 655
QY 440 GREFGLPNTSYGKLSVNOLETNSVDMREVMNMIIOGFQWNSKSVTTADTMLOKALE 499
DB 656 QALIGAANTGRRGSISGSKLESSNVLDSRLTNLVIVQRFQANSKAVTTSDOILNTLNL 715
QY 500 LKR 502
DB 716 LKQ 718

RESULT 29
Q92DV7
ID Q92DV7 PRELIMINARY; PRT; 411 AA.
AC Q92DV7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Lmc0697 protein.
GN OrderedLocusNames=lmc0697;
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RC MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstrek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
EMBL; AL591976; CAC98775.1; -
DR PIR; A11161; A11161.
```



|                                                                    |                                                                      |                                                            |              |      |     |     |
|--------------------------------------------------------------------|----------------------------------------------------------------------|------------------------------------------------------------|--------------|------|-----|-----|
| QX                                                                 | NCBI_TaxID=1642;                                                     | AA03514                                                    | PRELIMINARY; | PRT; | 411 | AA. |
| RN                                                                 | [1]                                                                  | AA03514                                                    |              |      |     |     |
| RP                                                                 | SEQUENCE FROM N.A.                                                   | AA03514                                                    |              |      |     |     |
| RC                                                                 | STRAIN=CLIP 11262 / Serovar 6a;                                      | AA03514                                                    |              |      |     |     |
| RX                                                                 | MEDLINE=21537279; PubMed=1169669;                                    | AA03514                                                    |              |      |     |     |
| RA                                                                 | Glaser P., Frangeul L., Buchrieser C., Rueniok C., Amend A.,         | AA03514                                                    |              |      |     |     |
| RA                                                                 | Baquerio F., Berche P., Bloeker H., Brandt P., Chakraborty T.,       | AA03514                                                    |              |      |     |     |
| RA                                                                 | Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,          | AA03514                                                    |              |      |     |     |
| RA                                                                 | Dannan E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., | AA03514                                                    |              |      |     |     |
| RA                                                                 | Entian K.-D., Fahi H., Garcia-del Portillo F., Garrido P.,           | AA03514                                                    |              |      |     |     |
| RA                                                                 | Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., | AA03514                                                    |              |      |     |     |
| RA                                                                 | Jones L.-M., Karst U., Kref J., Kuhn M., Kunst F., Kurapat G.,       | AA03514                                                    |              |      |     |     |
| RA                                                                 | Madueno E., Maitournam A., Maca Vicente J., Ng E., Medjari H.,       | AA03514                                                    |              |      |     |     |
| RA                                                                 | Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  | AA03514                                                    |              |      |     |     |
| RA                                                                 | Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,              | AA03514                                                    |              |      |     |     |
| RA                                                                 | Vogel B., J.-A., Voss H., Wehland J., Cossart P.;                    | AA03514                                                    |              |      |     |     |
| RT                                                                 | "Comparative genomics of <i>Listeria</i> species.";                  | AA03514                                                    |              |      |     |     |
| RL                                                                 | Science 294:849-852(2001)                                            | AA03514                                                    |              |      |     |     |
| RL                                                                 | EMBL; AL596166; CAC95937.1; -;                                       | AA03514                                                    |              |      |     |     |
| DR                                                                 | ListiList; LIN010705; -;                                             | AA03514                                                    |              |      |     |     |
| DR                                                                 | InterPro; IPR010930; DUF1078.                                        | AA03514                                                    |              |      |     |     |
| DR                                                                 | InterPro; IPR011491; FlaeE.                                          | AA03514                                                    |              |      |     |     |
| DR                                                                 | InterPro; IPR01444; Flag_bb_rod.                                     | AA03514                                                    |              |      |     |     |
| DR                                                                 | Pfam; PF06429; DUF1078; 1.                                           | AA03514                                                    |              |      |     |     |
| DR                                                                 | Pfam; PF07559; FlaeE; 1.                                             | AA03514                                                    |              |      |     |     |
| DR                                                                 | Pfam; PF04660; Flg_bb_rod; 1.                                        | AA03514                                                    |              |      |     |     |
| DR                                                                 | PROSITE; PS00588; FLAGELLA_BB_ROD; 1.                                | AA03514                                                    |              |      |     |     |
| KW                                                                 | Complete proteome.                                                   | AA03514                                                    |              |      |     |     |
| SQ                                                                 | SEQUENCE 411 AA; 42836 MW; 3272D338BB4E1A7C CRC64;                   | AA03514                                                    |              |      |     |     |
| Query Match 18.7%; Score 484.5; DB 2; Length 411;                  |                                                                      |                                                            |              |      |     |     |
| Best Local Similarity 27.0%; Pred. No. 1.9e-21;                    |                                                                      |                                                            |              |      |     |     |
| Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 11; |                                                                      |                                                            |              |      |     |     |
| QY                                                                 | 1                                                                    | MMGSLFICATGKTHSTGLTVSNANANTIGYKQOQVDFQDLPFSD----           | LAIGSTGS     | 56   |     |     |
| DB                                                                 | 1                                                                    | MMQMTYTAISGMNAFOQALSVTSNNANANTIGYKQSVFNDLLYQNTGMSVAGGLYAG  | 60           |      |     |     |
| QY                                                                 | 57                                                                   | QGNPQAGMAGVSVRTITGCAPEPNSVTDLAIGKGFQV--TLEDKVHYTRAGNER     | 114          |      |     |     |
| DB                                                                 | 61                                                                   | TNPMFSGSGKIGAILTDYTAGSPSTGRNKDAALQGRGFFIAGDNAGGNIVYTRDGEFA | 120          |      |     |     |
| QY                                                                 | 115                                                                  | FTQDGLNDPSPGFTLMSGRISNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNL  | 174          |      |     |     |
| DB                                                                 | 121                                                                  | VSDNNYLTQQGKYVMGYATDKNGVNLGNLQPIQLPLNSAIPGEATKNGSLSGNIPLDW | 180          |      |     |     |
| QY                                                                 | 175                                                                  | GDSTDKTQSEANPYFALLESWKNGTTPISTNSYSAQPMRVYDQOQNSHDITVYFDGA- | 233          |      |     |     |
| DB                                                                 | 181                                                                  | GEK-DTISSE-----LSVYDNAGGKHKLQVNMKAAT                       | 210          |      |     |     |
| QY                                                                 | 234                                                                  | PSSTGKTFEYLVMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK  | 293          |      |     |     |
| DB                                                                 | 211                                                                  | PDASGNVSYEYEQM-----DGKALTPP-----VTGTLNNAQGLTNPDA-----      | 251          |      |     |     |
| QY                                                                 | 294                                                                  | DLNMQAPLVLNGLPQFSANFVAGIQPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLP | 353          |      |     |     |
| DB                                                                 | 252                                                                  | -LKNIOINSTVNG-----KQVNM-----GLNLSGLT                       | 276          |      |     |     |
| QY                                                                 | 354                                                                  | SNMPTQSSGNSTARNSSSTRYSQDGYPOGLVDVITITSEGLKQKYSNQVDFYNI     | 413          |      |     |     |
| DB                                                                 | 277                                                                  | NY-----GTNQVPSPTSDGKAATVKDYAVTDSGYIAVSYSDGTVIPVAQL         | 322          |      |     |     |
| QY                                                                 | 414                                                                  | PLARFTSEDLREGNNHYSATLDSGGPFGLPFTSNYKLSVNOLETNSVDMGRMYNM    | 473          |      |     |     |
| DB                                                                 | 323                                                                  | AVATFSEDGLVKMGNGEYVPGVSSGDVYGVAGQAGGAGGSGSLEGSNVDLSREFVNL  | 382          |      |     |     |
| QY                                                                 | 474                                                                  | IIIQGFQNMKSVTTADTMLOKALEL                                  | 500          |      |     |     |
| DB                                                                 | 383                                                                  | MTYQSGFQGNKTVIRVADDMQKQIVNL                                | 409          |      |     |     |

ID Q7UCX0 PRELIMINARY; PRT; 402 AA.  
AC Q7UCX0; 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Flagellar hook protein FlgE.  
GN Name=flgE; OrderedLocusNames=S1160;  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2457T;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T."  
RL Infect. Immun. 71:2775-2786(2003).  
DR EMBL; A016981; AAP16590.1; -  
DR GO; GO:0003288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; Flab.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; Flab; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
KW Flagellum.  
SQ SEQUENCE 402 AA; 42051 MW; E13798711BD7B22F CRC64;

Query Match 18.7%; Score 483.5; DB 2; Length 402;  
Best Local Similarity 28.4%; Pred. No. 2.1e-21;  
Matches 141; Conservative 77; Mismatches 173; Indels 105; Gaps 15;  
QY 10 TGMKTHSTGLGVSNINANTIGYKQQVVDLFSQDLAIGSTGSGPQAGNQAQVG 69  
DB 8 SGLNAAATNLVDIGNANSATYGFKSTASPADMFA-----GS-----KVGLGVKVA 55  
QY 70 SVRTITQCAFPGNSVTDLAIGKGFQ-VTLEDKVHYTRAGNFRFTQDGLNDPSGFT 128  
DB 56 GITQDFTDGTNTTGRGLDVAISQNGFRLVDSNGSVFYSRNGQFKLDENRNLNTQGLQ 115  
QY 129 LMGSRISSNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSPTKTOSEANPY 188  
DB 116 LTGYPTGTPPTIQOAGNPTNISPTLMA---AKTTTASMQINL-NSSDPLPT----- 166  
QY 189 FALLESWKNGTPPTISTN---YSYAQPMRVYDQGNSHDITVYFDGAPSSGTSTFEYL 245  
DB 167 -----VTPFSASNADSYNKKGSVTVFDSQGNADMSVYF-----VKTGNNWQ-- 209  
QY 246 VAMNPSDGSASGTDAGLLMSGTMTFSSNGELKNMTAFTPTGSAKDLNANQAPLNV 305  
DB 210 VYTQSSDPSNSTAKT-----ATTLEFNAGTL--VDGAMANNIATGAINGAEP 256  
QY 306 GLPQPSANFVGAGIQLTDFGKISQQNWAGAPASAAAIGTDIGKLPMMPIQTSSGNS 365  
DB 257 ---TPSLGLFNS-----MQQN-----TGANNI 275  
QY 366 TARNSSSTRYSODGYPQGDVVDVITSEKLOGKYSNSQVDFYNIPLARFSEDLR 425  
DB 276 VAT-----TQNGYKPGDLVSVQINDGTVVGNNSQEQTLGQVLNANFANNEGLA 326  
QY 426 REGNNHYSATLDSGPFGLPFTSNYKGLSVNQLSTNSVNDMSRENMVNIITQRFQVNSK 485  
DB 327 SEGDVNSATQSSGVALLGTAAGTGNFGTLTNGALEASNVLDLSKELVNNIQAQRNYSNAQ 386

QY 486 SVTTADTMLOKALELK 501  
DB 387 TIKTODQILNTRVNL 402  
RESULT 34  
Q83RT3 PRELIMINARY; PRT; 642 AA.  
ID Q83RT3  
AC Q83RT3;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Flagellar biosynthesis, hook protein.  
GN Name=flgB; OrderedLocusNames=SF1080;  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=301 / Serotype 2a;  
RX MEDLINE=22272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,  
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
through comparison with genomes of Escherichia coli K12 and O157."  
RL Nucleic Acids Res. 30:4432-4441(2002).  
DR EMBL; A015135; AAN42702.1; -  
DR GO; GO:0002888; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; Flab.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR InterPro; IPR005648; FlgB.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; Flab; 1.  
DR Pfam; PF03963; FlgD; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
KW Complete proteome; Flagellum.  
SQ SEQUENCE 642 AA; 66632 MW; 64278E0207F67D34 CRC64;

Query Match 18.7%; Score 483.5; DB 2; Length 642;  
Best Local Similarity 28.4%; Pred. No. 4e-21;  
Matches 141; Conservative 77; Mismatches 173; Indels 105; Gaps 15;  
QY 10 TGMKTHSTGLGVSNINANTIGYKQQVVDLFSQDLAIGSTGSGPQAGNQAQVG 69  
DB 248 SGLNAAATNLVDIGNANSATYGFKSTASPADMFA-----GS-----KVGLGVKVA 295  
QY 70 SVRTITQCAFPGNSVTDLAIGKGFQ-VTLEDKVHYTRAGNFRFTQDGLNDPSGFT 128  
DB 296 GITQDFTDGTNTTGRGLDVAISQNGFRLVDSNGSVFYSRNGQFKLDENRNLNTQGLQ 355  
QY 129 LMGSRISSNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSPTKTOSEANPY 188  
DB 356 LTGYPTGTPPTIQOAGNPTNISPTLMA---AKTTTASMQINL-NSSDPLPT----- 406  
QY 189 FALLESWKNGTPPTISTN---YSYAQPMRVYDQGNSHDITVYFDGAPSSGTSTFEYL 245  
DB 407 -----VTPFSASNADSYNKKGSVTVFDSQGNADMSVYF-----VKTGNNWQ-- 449  
QY 246 VAMNPSDGSASGTDAGLLMSGTMTFSSNGELKNMTAFTPTGSAKDLNANQAPLNV 305  
DB 450 VYTQSSDPSNSTAKT-----ATTLEFNAGTL--VDGAMANNIATGAINGAEP 496  
QY 306 GLPQPSANFVGAGIQLTDFGKISQQNWAGAPASAAAIGTDIGKLPMMPIQTSSGNS 365

```
Db 497 ---TFSLFLNS-----MQQN-----TCANNI 515
Qy 366 TARGSSSTRYSDGYPQGLVDVTTITSEKLGKYSNQVDFYFNPIPLARTSEDDL 425
Db 516 VAT-----TONGYKPGDLVSYQINDDGTVYGNNSQEQTLGLQVILANFANNEGLA 566
Qy 426 REGNNHYSATLDGSGPFGPLPGTSNYGKLSVNOLETSNVDMSREVMNMIIQRFQWNSK 485
Db 567 SEGDNVWSATQSSVALLGTAGTCNFGTLITNGALEASNVLDLSKELVSMIVAQRYKSNQA 626
Qy 486 SVTTADTMLOKALELK 501
Db 627 TIKTQDQILNTRVNL 642

RESULT 35
Q6D6H6 PRELIMINARY; PRT; 405 AA.
ID Q6D6H6
AC Q6D6H6
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE FLAGELLAR hook protein.
GN Name=flgE; Synonyms=flak; ORFNames=ECAL705;
OS Erwinia carotovora subsp. atroseptica SCRI1043.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=218491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI1043;
RA Bell K.S., Sebalhia M., Pritchard L., Holden M., Hyman L.J.,
RA Holeva M.C., Thomson N.R., Bentley S.D., Churcher C., Mungall K.,
RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA Ormond G.P.C., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA Salmund G.P.C., Birch P.R.J., Barrell B.G., Parkhill J., Toth I.K.,
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; BX950851; CAG74611.1;
KW Flagellum.
SQ SEQUENCE 405 AA; 41910 MW; 687FB7C32380814B CRC64;

Query Match 18.6%; Score 482; DB 2; Length 405;
Best Local Similarity 28.6%; Pred. No. 2.6e-21;
Matches 142; Conservative 75; Mismatches 17; Indels 104; Gaps 16;

Qy 10 TGMKTHSTGLTVSNNIANANTIGYKQQVVFQDLFSODLAIGSTGSGPNQAGMGQV 69
Db 8 SGLNRAASNLVDVGNNIANSATVGFKSGNVTFADNFA-----GS-----KVGMGVKA 55
Qy 70 SVRTITFOGAFEPGNSVTDLAIGKGFQVLTED---KVHYTRAGNFRFTQDGLNDPSG 126
Db 56 SVLQDFGNGTVTSSSRDLDAISGGGFYR--LQDTNGSTYYSRNGQFMLNGRNIWN-AQG 112
Qy 127 FTLMGSRISNNPNIKKETLEPIOLDNDPTVAKSPAKTSTALNAVMLGSDTDKTOGEAN 186
Db 113 MOLTGYFVAGTPPTVQTGADVPVLTPDGMLAS--QTSVA-SIKANL-KSSDSVPINA- 167
Qy 187 PYFALLESWKNGTTPPISTSNYSYQAQPMRVYDQGNSHDITVYFDGAPSSGSKTFEVLV 246
Db 168 -----W-----ATTPGAEGTNSKTALTITDYSQGNVHNFTLYFVKYANNTW-QTYAKDD 215
Qy 247 AMNPEDCSAASGDSAGLLMGTTTFSSNGELKN--MTATPTGSAKDLNAQAPPLV 304
Db 216 SINPATYONA-----GTLNFANGALSTGATAHTP----- 246
Qy 305 NGLPOFSANFVAGIQPLTLDFGIKXQQNMWAGAPASAAAIGTDIGKLPMMPIQTSNGN 364
Db 247 -----FTLNLG-----TNGAANGT-----FTLNLG 268
Qy 365 STARNSSSTRYSQDGPQGLVDVTTITSEKLGKYSNQVDFYFNPIPLARTSEDDL 424
Db 269 SVQONTESYKSTQNTQNGAPGSLTGFAINDDGTIEGYSNGOKQALQILLASFANPEGL 328
```

```
Qy 425 REGNNHYSATLDGSGPFGPLPGTSNYGKLSVNOLETSNVDMSREVMNMIIQRFQWNS 484
Db 329 SPEGDNVWSATQSSVALLGTAGTCNFGTLITNGALEASNVLDLSKELVSMIVAQRYKSNQA 388
Qy 485 KSVTTADTMLOKALELK 501
Db 389 QTIKTQDSILOTLVSLR 405

RESULT 36
Q7M8B3 PRELIMINARY; PRT; 716 AA.
ID Q7M8B3
AC Q7M8B3
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE FLAGELLAR HOOK PROTEIN.
GN Name=flgE; OrderedLocusNames=WS1758;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;
RX MEDLINE=22882897; PubMed=14500908;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RA "Complete genome sequence and analysis of Wolinella succinogenes.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695 (2003).
RL EMBL; BX571661; CAB0780.1;
DR GO; GO:0003288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0001598; F:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR010810; Flagellin_IN.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07359; Flae; 1.
DR Pfam; PF07196; Flagellin_IN; 1.
DR Pfam; PF00460; Flg_bb rod; 1.
DR Complete proteome; Flagellum.
SQ SEQUENCE 716 AA; 74684 MW; 28C478F91D30C605 CRC64;
```

```
Query Match 18.6%; Score 482; DB 2; Length 716;
Best Local Similarity 24.8%; Pred. No. 5.8e-21;
Matches 183; Conservative 79; Mismatches 217; Indels 260; Gaps 21;

Qy 1 MMSLFIATGKTHSTGLTVSNNIANANTIGYKQQVVFQDLFSODLA-----GSTG 55
Db 1 MLRSLSWGSVGMQAHQVALDVEGNINANTYNTTGFKYSRANFSDMLGSQVNRATSPYGGLG 60
Qy 56 SQGPNQAGMGQVGSVRTITFOGAFEPGNSVTDLAIGKGFQVLTED---KVHYTRAGNF 113
Db 61 GQNDYSIGLGTSINSTTKIFGQSGIQDTTNKWDLAIGDGGFFIVSGNGGRTNAYTRDGAF 120
Qy 114 RFTQDGLNDPSGFTLMG-SRISNNPNIK-----KETLEPIOLDNDPTVAKSPAK 163
Db 121 GPDAAAGNMVNNAGYIVQGWTRDLNLSGTSYSDALYNVVDVTPVPSIGIKIEPKQW-IPAK 179
Qy 164 TSTALNAVNL--GDSTDK-----TQSEANPYFALLSWKNGKNGTTPPISTSNYSYQA 212
Db 180 ATTQVNLDAWLTAGDITDKLGCWYALDSVTAADGIAARYDSAGN-----KIQMAE 231
Qy 213 PMRV-YDQGN-----SHDITV----- 228
Db 232 DMGVLFNASNALKSEGGQGVVSYQATQAVAVATTGTLTNGTITFTNDSTISGV 291
Qy 229 -----YFDGAPSSGSKTF-----EYLAVNMPSEDGSA-----ASG----- 259
```

Db 292 SSLVAQAQNAKSGTGEATGQGLRLVNDNSLDGNASKQNVIIITASGALANFTAA 351  
QY 260 -----TDSAGLMSGTMT 272  
Db 352 DNSITAFRAYTASDADSTSLRFTEDLRLALLODANNIKHGGYVDSGTNASVKVT 411  
QY 273 FSNGBELK-----NMFTPTGSAATKDLNA-----297  
Db 412 INKTGMFEILNODGDTTTCNLISLTVSSYDTNVTISNVLFRKSAKGLNTGILVEGSSSTT 471  
QY 298 -----W-----OPALVNLGPQFSANF 314  
Db 472 SASLMAAKHTATDIDVSLCNKHTLTVTFRKVGPKQKSFSLHVPFATFVNGSGERPNYF 531  
QY 315 VG-----AGIQPTLDFGKSKQNMWAGAPASAAAIGTDIGKLPNMPIQTSSG 363  
Db 532 EGRVTFGSDGLTGMNPTTIOFNPKS-----GA-SSPORIDLDFG-----VSGTFQ 577  
QY 364 NSTARNGSSSTRYSODGYPQGLVDVITTSSEKLGKYSKSOVYDFYNIPLABETSEDG 423  
Db 578 GLTSTDKSATGNIVYQNGQSGVLEMDRFDNSGVLTGFSNGKOLALAOVAIASFTNNGG 637  
QY 424 LRREGNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMRSVMNMIIIOQGFQNM 483  
Db 638 LQAGSNLFSQTANSGEPPVGTAGSGGRGKISPSALEMSNVDSLSRLTQLIIVQRGFOAN 697  
QY 484 SKSVTTADTMLOKALELKR 502  
Db 698 SKVTTSDDQILNTLLQLKQ 716

## RESULT 37

Q8P9B6 PRELIMINARY; PRT; 407 AA.  
ID Q8P9B6  
AC Q8P9B6  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Flagellar biosynthesis, hook protein.  
GN Name=figE; OrderedLocusNames=XCC1949;  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=340;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33913 / NCPPB 528;  
RA MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Mencia C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463(2002).  
DR EMBL; AE012295; AAM41238.1;  
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.

DR InterPro; IPR011491; Flae.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; Flae; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_rod; 1.  
KW Complete proteome; Flagellum.  
SQ SEQUENCE 407 AA; 42710 MW; 307BE1287BE2B079 CRC64;  
Query Match 18.6%; Score 481; DB 2; Length 407;  
Best Local Similarity 27.2%; Pred. No. 3e-21;  
Matches 137; Conservative 77; Mismatches 174; Indels 116; Gaps 13;  
QY 10 TGMKTHSTGTCVSNNIANANTIGYQQQVQDLPQDLPSODLAIGSTGSCQPNQAGKAQVIG 69  
Db 8 SGINAANADLVNTSNNIANVNTTGFKESRAEPADMF-QSTSYGLS-----RNAVSGSVRS 62  
QY 70 SVRTITQGAPEPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLNDPSGFTL 129  
Db 63 NVAQQFSQGNIDTGRSLDLAISGEFFAVSSNGAKWYTRAGNFQDANGYVINPQGAEL 122  
QY 130 MGRISNNPN---IKKETLEPIQLDNDPTVAKSPAKTSTALNAVNL-GDSTDKTQSEA 185  
Db 123 --QVFAPNPSGNGFDVGRSLDLQLLTD-----SPPKSTSTVNLAFPLFGNASAPTVPF 175  
QY 186 NPYFALLESMKNGTTPPISTSNVSYAQ-PMRVYDQGNSHDITVYFDGAPSPSTGSKTPEY 244  
Db 176 N-----PADDKTYSHSTGGINVDYSLGVSHVQTSYF-----206  
QY 245 LVAMNPSE-----DGSAASTDSAGLLSGTMTPTFSSNGELKNMTAFTPTGSATKDLNA 297  
Db 207 VKTANPNQVHNYVDGSAVGAPTTLQFSDTGLTTPANG-IVAMDPTPTS-----256  
QY 298 WQAPLVNGLPQFSANFVAGIQTPLTDFGKSKQNMWAGAPASAAAIGTDIGKLPSPMP 357  
Db 257 -----TGAGV-----LS 263  
QY 358 ICTSSGNSTARNGSSSTRYSODGYPQGLVDVITTSSEKLGKYSNSQVDFYNIPLAR 417  
Db 264 MQLNVSGSTQYGEAPALRDTQDGYASGLNEISDTSQVWFARYSNGADKALGQVALNS 323  
QY 418 FTSEDGLRREGNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMRSVMNMIIIO 477  
Db 324 FVNPQGLQSGNNMWAESYTSGAARVGAPDTSDLQIQSGSLEASTVDLTQELVNMVIAQ 383  
QY 478 RGFQNMKSQVTTADTMLOKALELK 501  
Db 384 RNFQANSQWISTQDQVQVTTIINR 407  
RESULT 38  
Q8PL23 PRELIMINARY; PRT; 407 AA.  
ID Q8PL23  
AC Q8PL23  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Flagellar biosynthesis hook protein.  
GN Name=figE; OrderedLocusNames=XAC1983;  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RA MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,  
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Mencia C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463(2002).  
DR EMBL; AE012295; AAM41238.1;  
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tezi S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.,  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities".  
RL Nature 417:459-463(2002).  
DR EMBL; AB011833; AM36845.1; --  
DR GO; GO:000288; C:flagellum (sensu Bacteria); IEA.  
DR GO; GO:0003774; F:motor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
DR InterPro; IPR010930; DUF1078.  
DR InterPro; IPR011491; Flae.  
DR InterPro; IPR001444; Flag\_bb\_rod.  
DR Pfam; PF06429; DUF1078; 1.  
DR Pfam; PF07559; Flae; 1.  
DR Pfam; PF00460; Flg\_bb\_rod; 1.  
DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
DR Complete proteome; Flagellum.  
KW SEQUENCE 407 AA; 42762 MW; 57E2E732B6069F9C CRC64;  
SQ  
Query Match 18.6%; Score 481; DB 2; Length 407;  
Best Local Similarity 27.0%; Pred. No. 3e-21;  
Matches 134; Conservative 80; Mismatches 181; Indels 102; Gaps 12;  
QY 10 TGMKTHSTGLGVSNANANTIGYKQOVVFDLFSQDLAIGTSQSGNQAGMAQVG 69  
DB 8 SCINAANADLVNTSNANVNTGPKESRAEFDMF-QSTSYGLS-----NNAVSGVRVS 62  
QY 70 SVRTFTTQGAPEPNSVTDLAIGKGFQVLEDKVHYTRAGNRFRTQDGLNDPSGFTL 129  
DB 63 NVAQQFSQGNIDPTGRSLDVLAVSGEGFTVSSNGAKMYTRAGNRFQDANGVYVNPQARL 122  
QY 130 MGRSISNNPN---IKKETLEPTQLDFNDPTVAKSPAKTSTALNAVVL-GDSDTKTOSEA 185  
DB 123 ---QVFAFPNSGNGPDVGRSLDQLLTDD-----SPPKSTSVNLAFPLGNATAPVTP? 175  
QY 186 NPYFALLSKWNGHTPPSTSNYSYQ-PMRVYQCGNSHDITVYFGAPSSGSKTFEY 244  
DB 176 N-----PADDKTYSHSTGGINVDYSLGSHVQTSYFVKTANPNQWQVINY 220  
QY 245 LVAMNPSDEGSAAGTDSAGLLMSGTMTFFSNGELKNMTAFTPTGSKATKOLNAWQAPLV 304  
DB 221 V-----DGAAGVAPTLQSDTGALTTPANGIIA-MDPTFS-----256  
QY 305 NGLPQFSANFVAGIQPTLDFGIKSSQQWAGAPASAAIGTDIGKLPENMPIQTSSGN 364  
DB 257 -----TCAGV-----LNNQLNVTG 270  
QY 365 STARNGSSSTRYSODGPQGDVDTVTTSBKLGKYSNSQVDFNYNPLARFTSDGL 424  
DB 271 STQYGEAFALRDTRODGYASGKLNIEISIDTSGVFPARYSGADKPLQVALSFFVNPGL 330  
QY 425 REGNNHYATLDGGPFGFLPGTSNYKLSVNOLETNSVNDMSHEMNMIIIOGFGQMS 484  
DB 331 QSCGNMMAESTYGAARTGAPDTSGLQIESGSLASTVDTLPOLVNMIVAGRNFOANS 390  
QY 485 KSVTTADTWMLOKALELK 501  
DB 391 QMISTQDQVQTIIINR 407  
RESULT 39  
FLGE HELPY  
ID FLGE HELPY STANDARD; PRT; 718 AA.  
AC PS0610;  
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Flagellar hook protein flgs.  
GN Name=flgs; OrderedLocNames=HP0870;  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=210;  
RN (1)  
RP SEQUENCE FROM N.A., SEQUENCE OF 1-25, AND CHARACTERIZATION.  
RC STRAIN=CCUG 17874 / NCIC 11638, and CCUG 951;  
RX PubMed=7891557;  
RA O'Toole P.W., Kostzynska M., Trust T.J.;  
RT "Non-motile mutants of Helicobacter pylori and Helicobacter mustelae  
RT defective in flagellar hook production.";  
RL Mol. Microbiol. 14:691-703(1994).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695 / ATCC 700392;  
RX MEDLINE=97394467; PubMed=9252185; DOI=10.1038/41483;  
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S.R.,  
RA Dougherty B.A., Nelson K.E., Quackenbush J., Zhou L., Kirkness E.F.,  
RA Peterson S.N., Loftus B.J., Richardson D.L., Hodson R.J., Khaliq H.G.,  
RA Glöckle A., McKenney K., Fitzgerald L.M., Lee N., Adams M.D.,  
RA Hickey E.K., Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D.,  
RA Kelley J.M., Cotton M.D., Weidman J.F., Fujii C., Bowman C.,  
RA Wathney L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D.,  
RA Smith H.O., Fraser C.M., Venter J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
RT pylori.";  
RL Nature 388:539-547(1997).  
CC -!- FUNCTION: A flexible structure which links the flagellar filament  
CC to the drive apparatus in the basal body. Absence of the gene  
CC leads to absence of the hook protein, lack of the flagellar  
CC filament and thus loss of motility. Approximately wild-type levels  
CC of the flagellar subunits are still produced and accumulate mostly  
CC in the cytosol.  
CC -!- SUBCELLULAR LOCATION: Envelope.  
CC -!- MISCELLANEOUS: The N-terminal amino acid sequence was from strain  
CC CCUG 951; the DNA sequence was from strain CCUG 17874. Disruption  
CC and subcellular localization experiments were performed for both  
CC strains. The apparent pI for the hook protein was 5.1.  
CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; U09549; AAA92803.1; --  
CC PIR; P64628; AAD07911.1; --  
CC TIGR; HP0870; --  
CC InterPro; IPR010930; DUF1078.  
CC InterPro; IPR010810; Flagellin\_IN.  
CC Pfam; PF06429; DUF1078; 1.  
CC Pfam; PF07196; Flagellin\_IN; 1.  
CC Pfam; PF00460; Flg\_bb\_rod; 1.  
CC PROSITE; PS00588; FLAGELLA\_BB\_ROD; FALSE NEG.  
KW Complete proteome; Direct protein sequencing; Flagellum.  
FT CONFLICT 106 106 S -> N (in Ref. 1).  
FT CONFLICT 237 237 C -> Y (in Ref. 1).  
FT CONFLICT 259 259 A -> P (in Ref. 1).  
FT CONFLICT 373 374 DI -> NS (in Ref. 1).  
FT CONFLICT 413 414 SV -> TI (in Ref. 1).  
FT CONFLICT 509 509 E -> D (in Ref. 1).  
SQ SEQUENCE 718 AA; 76206 MW; F36EB50684AD1333 CRC64;

|                           |                                                                      |                                                                  |
|---------------------------|----------------------------------------------------------------------|------------------------------------------------------------------|
| Query Match               |                                                                      | 18.6%; Score 481; DB 1; Length 718;                              |
| Best Local Similarity     |                                                                      | 24.8%; Pred. No. 6.7e-21;                                        |
| Matches 178; Conservative |                                                                      | 96; Mismatches 223; Indels 226; Gaps 21;                         |
| QY                        | 1                                                                    | MMGSLFIGATGKTHSTGLTGVSNINIANANTIGYKQQQVVFQDLFSQDLAIGS-----TGS 56 |
| DB                        | 1                                                                    | MLRSLWGVNGVMAHQALDIESNNIANVNTGFKYSRASFDVMDLSQVKTATAPYKNGL 60     |
| QY                        | 57                                                                   | QGEN--QAGNQAGVSVRIFFQGAPEPGNSVTDLAIGKGFQVTLDEKV--HYTRAGN 112     |
| DB                        | 61                                                                   | AGNDPSVLGTVGVDATTKIFSGNQNTDVTDLAIQDGFIIISPRGITRNFTRDGE 120       |
| QY                        | 113                                                                  | FRFTQDGLNDPSGFTLM-----GSRISNNPIKKTLEPIQLQDFNPVAKSPAK 163         |
| DB                        | 121                                                                  | FUFDSQGLVTTGGLVVGWVWNGSDTGNGKSDTDALKVDNTGPLENIRIDPGWV-MPAR 179   |
| QY                        | 164                                                                  | TST--ALNAVNLGSDTKTQSEANPYFALLESWK--GNGTPPISTSNYSYQA----- 212     |
| DB                        | 180                                                                  | ASNRISMRLNAGRHADQTA--IFALDSSAKTPSDGINPVYDSGTNLQAQVAEDMGS 235     |
| QY                        | 213                                                                  | -----PMRVVDQDQ-----NSHDITVYFDGAP-----SSTGSKTFEY 244              |
| DB                        | 236                                                                  | LCNEDGDALLINENQIWIYSKAKVVDIILPSAENSTLELNGVKISFTNDSAVRTSSL 295    |
| QY                        | 245                                                                  | LVAMPSEDSGAASG-----TDSAGLLMSGTMTFFSNGELKN----- 281               |
| DB                        | 296                                                                  | VAAKNAINAVKSGTGEAYLDGKQLRLNTNELDGEKLNIVVTQAGTGAFAFLDGDK 355      |
| QY                        | 282                                                                  | -MTAF-----TPGSAKDLNANQAPLVNGLPOPSANFV-----GAGIQPLTL 324          |
| DB                        | 356                                                                  | DVTAPEKYSYTHSISNADIGOFRTTEDRLALIQHDANIVDPSLADNYQDSAAISGVSN 415   |
| QY                        | 325                                                                  | DFG-----IKSQQNM----- 343                                         |
| DB                        | 416                                                                  | QVGMFEINKNKNIKENLIFVSGYSSDSVTNNVLKNAWKGLNTASLIEGASASS 475        |
| QY                        | 344                                                                  | -----AIGT----- 347                                               |
| DB                        | 476                                                                  | KPTHATHATSIDVLSLGTKHAMRIEYRSGAENFRVIVPEGELVGGSAARPNVEG 535       |
| QY                        | 348                                                                  | -----DIGKLPSMMP----- 379                                         |
| DB                        | 536                                                                  | GRHFNNDGSLAGWPPPLQFPNGADAPORINLAFSGSGFDLTSVDKISVTAIEQ 595        |
| QY                        | 380                                                                  | DYPOGLDVTITISEGKLGKYSNQVDPYNIPLARFTSEDLRREGNNHYSATLDG 439        |
| DB                        | 596                                                                  | NGYQAGDLMDVRFDSGVLGAFSGNRTLALAQAVALANFANDAGLQALCGNVFSGTNSG 655   |
| QY                        | 440                                                                  | GPEGLPCTSNYKGLSVNOLSTSNVDSREVMNIIIORGFOMNKSVTATDMLQKALE 499      |
| DB                        | 656                                                                  | QALIGAANTGRGSGISGLESSNVDLSRLTNLIVVQRFQANSKAVTTSDQILNTLN 715      |
| QY                        | 500                                                                  | LKR 502                                                          |
| DB                        | 716                                                                  | LKQ 718                                                          |
| RESULT 40                 |                                                                      |                                                                  |
| ID                        | Q7VFH2                                                               | PRELIMINARY; PRT; 718 AA.                                        |
| AC                        | Q7VFH2;                                                              |                                                                  |
| DT                        | 01-OCT-2003 (Tremblrel. 25, Created)                                 |                                                                  |
| DT                        | 01-OCT-2003 (Tremblrel. 25, Last sequence update)                    |                                                                  |
| DT                        | 01-MAR-2004 (Tremblrel. 26, Last annotation update)                  |                                                                  |
| DE                        | Flagellar hook protein Figs.                                         |                                                                  |
| CN                        | Name=fige.1; OrderedLocusNames=HH1704;                               |                                                                  |
| CS                        | Helicobacter hepaticus.                                              |                                                                  |
| OC                        | Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; |                                                                  |
| OC                        | Helicobacteriaceae; Helicobacter.                                    |                                                                  |
| OX                        | NCBI_TaxID=32025;                                                    |                                                                  |
| RN                        | [1]                                                                  |                                                                  |
| RP                        | SEQUENCE FROM N.A.                                                   |                                                                  |
| Query Match               |                                                                      | 18.5%; Score 479; DB 2; Length 718;                              |
| Best Local Similarity     |                                                                      | 23.6%; Pred. No. 8.8e-21;                                        |
| Matches 176; Conservative |                                                                      | 90; Mismatches 209; Indels 270; Gaps 21;                         |
| QY                        | 1                                                                    | MMGSLFIGATGKTHSTGLTGVSNINIANANTIGYKQQQVVFQDLFSQDLAIGST----- 54   |
| DB                        | 1                                                                    | MLRSLWGVNGVMAHQALDIESNNIANVNTGFKYSRASFDVMDLSQVKTATAPYKNGL 60     |
| QY                        | 55                                                                   | SGSQPNQAGVQAVSVRTIFQGAPEPGNSVTDLAIGKGFQVTLDEKV--HYTRAGN 112      |
| DB                        | 61                                                                   | GQNDPSVLGTVGVDATTKVFSQNTQNTDVKTDVAIEGDDGFFIISPRGTHNTRNGE 120     |
| QY                        | 113                                                                  | FRFTQDGLNDPSGFTLMG-----SRISNNPIKKTLEPIQLQDFN 153                 |
| DB                        | 121                                                                  | FLFDANGNLVTTGGVVGWVWVPPLEAAESGTMDSDFDFRVDNTGPNVNIQIDP----- 174   |
| QY                        | 154                                                                  | DPTVAKSPAKTSTALNAVNLG-----DSTDKTQSEA----- 185                    |
| DB                        | 175                                                                  | -GMVMPARATKTTITURANLNAGRHDIDQDVAALDSTAQTAADGVVAIYDSRGVLTQVGE 233 |
| QY                        | 186                                                                  | -----NYPFALLE-----SWK-----GNGTPPISTN----- 207                    |
| DB                        | 234                                                                  | DLAGVLFNDGDAFALNENQIWLKYKTAARHETVTVNEVSTIGINGEKVSFNNSAITG 293    |
| QY                        | 208                                                                  | -----YSYAPQMRVYDQ----- 219                                       |
| DB                        | 294                                                                  | VSSIVAAQNAINSLEKDTGVNAVYDAGQLRIENRNMQDGGKVKNRITSGTGVQLQNFV 353   |
| QY                        | 220                                                                  | QGNSHDITVY-----PDGAPSTGS-KTEFYLAVM-----NPSE-----DGSAASTG 260     |
| DB                        | 354                                                                  | QG-EEDITAFRYRTKSEDASTTGQFRTTEDRLALIQDANWIKNPEKTYESTATVGV 412     |
| QY                        | 261                                                                  | -----DSAGLLMSGTMTFSSN----- 276                                   |
| DB                        | 413                                                                  | TINRWGMEFIANHDDADDEQRNLSLYVTSFSDVNTNNVLKFTWKALNTASLIEGGA 472     |
| QY                        | 277                                                                  | -----GELKNMT-----AFTPGSATKDLNANQAPLVNGLPOPSANFV 315              |
| DB                        | 473                                                                  | NTGKIVKATHATSVDI VDSLGSKINVRPEFWKTGDVWSPRAIVPEP-----AQFI 523     |
| QY                        | 316                                                                  | GAGIQPLTDFGIKSQQNMWAGAPASAAIGTDIGKLFPMWPIQTSNGS----- 365         |
| DB                        | 524                                                                  | GGS1-----TKENVPEGGRASFNSDGLSGMNPVLPQDPDKGSGKQPRLKLF 573          |
| QY                        | 366                                                                  | -----TARNGSSSTRYSQDGPQGLVDVTTITSEGKLGKYSNQVVDVFNIPAR 417         |
| DB                        | 574                                                                  | ANETFGGLTSVDKISYTSINQYQAGDLMDIRFDSNGSLGAFSNGRSIALAQAVALAN 633    |

Search completed: October 26, 2004, 09:09:00  
Job time : 206 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 31, 2004, 02:44:06 ; Search time 3026 Seconds  
(without alignments)

6045.190 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMSLFGATGKMTHTSTGLG.....NSKSVTTADTMLOKALELKR 502

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US10009823/runat\_26102004\_100129\_1617/app.query.fasta\_1.647  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALALIGN=200 -THR\_SCORES=0 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10009823 @CGN\_1\_1 2607 @runat\_26102004\_100129\_1617 -NCP=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT\_DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hcc:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gsel:  
9: gb\_gsel2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID         | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| C 1        | 294   | 11.4        | 729    | 8 BH370667 | BH370667 AG-ND-165 |
| C 2        | 272   | 10.5        | 891    | 8 B2577315 | B2577315 msh2_5350 |
| C 3        | 221.5 | 8.6         | 1319   | 8 B2556688 | B2556688 pasc1-60  |
| C 4        | 221   | 8.5         | 742    | 8 BH795484 | BH795484 23H03LL10 |
| C 5        | 214   | 8.3         | 1105   | 8 B2563024 | B2563024 pasc2-164 |
| C 6        | 209   | 8.1         | 752    | 7 CF891171 | CF891171 UI-CF-DUI |
| C 7        | 208.5 | 8.1         | 895    | 8 B2577405 | B2577405 msh2_5391 |
| C 8        | 206.5 | 8.0         | 565    | 6 CB065273 | CB065273 EST644954 |
| C 9        | 200.5 | 7.8         | 898    | 7 CO423013 | CO423013 GGEZHT102 |

|      |       |     |      |            |                    |
|------|-------|-----|------|------------|--------------------|
| C 10 | 200.5 | 7.8 | 1190 | 8 BZ557649 | BZ557649 pasc1-60  |
| C 11 | 200   | 7.7 | 849  | 9 CL660824 | CL660824 PRI0138a  |
| C 12 | 199.5 | 7.7 | 824  | 7 CO202677 | CO202677 Oa splbn  |
| C 13 | 197.5 | 7.6 | 758  | 7 CO421506 | CO421506 GGEZHT102 |
| C 14 | 197.5 | 7.6 | 761  | 1 AJ747468 | AJ747468 AJ747468  |
| C 15 | 197.5 | 7.6 | 766  | 7 CO423511 | CO423511 GGEZHT101 |
| C 16 | 190.5 | 7.4 | 870  | 7 CN823216 | CN823216 Oa splbn  |
| C 17 | 183   | 7.1 | 650  | 6 CB854175 | CB854175 UI-CF-DUI |
| C 18 | 181.5 | 7.0 | 661  | 7 CN582719 | CN582719 USDA-FP_1 |
| C 19 | 178   | 6.9 | 566  | 5 BP562788 | BP562788 BP562788  |
| C 20 | 175   | 6.8 | 436  | 6 CD119056 | CD119056 ME1-0052U |
| C 21 | 175   | 6.8 | 436  | 6 CD119187 | CD119187 ME1-0052U |
| C 22 | 172.5 | 6.7 | 532  | 1 AI057966 | AI057966 SMOVL3CAN |
| C 23 | 169   | 6.5 | 715  | 5 CD444654 | CD444654 EL01N0442 |
| C 24 | 168   | 6.5 | 692  | 7 CF886932 | CF886932 UI-CF-DUI |
| C 25 | 161.5 | 6.2 | 1158 | 8 BZ557676 | BZ557676 pasc1-60  |
| C 26 | 161.5 | 6.2 | 1208 | 8 BZ557650 | BZ557650 pasc1-60  |
| C 27 | 158.5 | 6.1 | 554  | 4 BI309317 | BI309317 EST530727 |
| C 28 | 154   | 6.0 | 747  | 7 CN822868 | CN822868 Oa splbn  |
| C 29 | 139   | 5.4 | 3057 | 3 AK044947 | AK044947 Mus muscu |
| C 30 | 138   | 5.3 | 374  | 8 AZ577776 | AZ577776 14f11 Sho |
| C 31 | 137.5 | 5.3 | 1590 | 9 AY412935 | AY412935 Homo sapi |
| C 32 | 135.5 | 5.2 | 4428 | 9 AY416869 | AY416869 Pan trogl |
| C 33 | 133.5 | 5.2 | 726  | 7 CN823464 | CN823464 Oa splbn  |
| C 34 | 133.5 | 5.2 | 4355 | 9 AY416870 | AY416870 Mus muscu |
| C 35 | 132   | 5.1 | 374  | 7 CN762396 | CN762396 IDOAAA4AB |
| C 36 | 131.5 | 5.1 | 348  | 4 BG857482 | BG857482 1024052B0 |
| C 37 | 131.5 | 5.1 | 1723 | 3 AK089973 | AK089973 Mus muscu |
| C 38 | 128.5 | 5.0 | 324  | 8 AZ577764 | AZ577764 14e10 Sho |
| C 39 | 128.5 | 5.0 | 4428 | 9 AY416868 | AY416868 Homo sapi |
| C 40 | 128   | 4.9 | 867  | 8 BZ577319 | BZ577319 msh2_5352 |
| C 41 | 127.5 | 4.9 | 527  | 6 CA581013 | CA581013 EST000688 |
| C 42 | 126.5 | 4.9 | 882  | 8 BZ559876 | BZ559876 pasc2-164 |
| C 43 | 124.5 | 4.8 | 2293 | 3 AK032791 | AK032791 Mus muscu |
| C 44 | 124.5 | 4.8 | 4722 | 3 AK029775 | AK029775 Mus muscu |
| C 45 | 123   | 4.8 | 4489 | 3 AK028459 | AK028459 Mus muscu |

#### ALIGNMENTS

#### RESULT 1

BH370667/c

LOCUS BH370667 729 bp DNA linear

DEFINITION AG-ND-165M21.TF ND-TAM Anopheles gambiae genomic clone

ACCESSION AG-ND-165M21, genomic survey sequence.

VERSION BH370667

KEYWORDS BH370667.1 GI:17316792

SOURCE Anopheles gambiae (African malaria mosquito)

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

1 (bases 1 to 729)

REFERENCE Hong, X.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J.,

Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B.,

Gardner, M.J. and Collins, F.H.

Construction of a BAC library and generation of BAC end

sequence-tagged connectors for genome sequencing of the African

malaria mosquito Anopheles gambiae

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE 22542063

PUBMED 12655398

COMMENT Other GSSs: AG-ND-165M21.TR

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

# FEATURES

source  
Location/Qualifiers

1..729  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-165M21"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

## ORIGIN

Alignment Scores:  
Pred. No.: 2,06e-20 Length: 729  
Score: 294.00 Matches: 729  
Percent Similarity: 46.84% Conservative: 38  
Best Local Similarity: 30.64% Mismatches: 77  
Query Match: 11.37% Indels: 48  
DB: 8 Gaps: 3

US-10-009-823A-1 (1-502) x BH370667 (1-729)

Qy 267 MetSerGlyThrMetThrPheSerSerAsnGlyLeuLeuLysAsnMetThrAlaPheThr 286  
Db 727 ATGCCCGGTACATGACGTTTCGACACCGGCAATCTGGCCAGCACCCACGACCAACAA 668  
Qy 287 ProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGly 306  
Db 667 GGCACCGCGCGGATTCAGATCTG----- 641  
Qy 307 LeuProGlnPheSerAlaAsnPhaValGlyAlaGlyIleGlnProLeuThrLeuAspPhe 326  
Db 640 ATCCCGATGAAGGCAAGAT-----GGCGCGCGGCACAGAACTTCACCTGGAATTC 587  
Qy 327 GlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGly 346  
Db 586 GCCAAGACATGCAGACGCGTC----- 563  
Qy 347 ThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThr 366  
Db 563 ----- 563  
Qy 367 AlaArgAsnGlySerSerThrArgArgTy-SerGlnAspGlyTyProGlnGlyAsp 386  
Db 562 -----AGCGCGGACTCCGTTCAGCAAGATGTCGCGAGGATGCTATGGCGGAGTAA 512  
Qy 387 LeuValAspValThrIleThrSerGlyLysLeuGlnGlyLysTySerAsnSerGln 406  
Db 511 TACACCACTTCAGATCAACACGAGCGGACCGTATGGGCATCTACTTCCACACGAGAG 452  
Qy 407 ValValAspPheTyThrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426  
Db 451 ACTCAGGTGTGGCGCAGATGCTATGGGCAACTTCTCCAAACCGGAGGCGCTGTCTCT 392  
Qy 427 GluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyGlyProGlnPheGlyLeuPro 446  
Db 391 CAGGGGATACCTTTGGGAGAAACCGCGGATCCCGGATCCGACCGATAGCTGTGGCG 332  
Qy 447 GlyThrSerAsnTyGlyLysLeuSerValAsnGlnLeuGlnThrSerAsnValAspMet 466  
Db 331 GGCAACGGGGTGTGGGCAAACTGACAGCGGCGGCTGGAAGCTCCAAACGCTCATCTG 272  
Qy 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486  
Db 271 AGCCAGAACTGGTGAACATGATCTGCGCACCAACGTAACCTACGTCGGAACGCCAACCC 212

Qy 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501  
Db 211 ATCAAAACGAGGACTCGATTCTGCAACGCTGTTAGCTGGCG 167

## RESULT 2

BZ577315 891 bp DNA linear GSS 17-DEC-2002  
LOCUS msh2\_5350.x1 msh Pseudomonas aeruginosa genomic clone msh2\_5350,  
DEFINITION genomic survey sequence.

ACCESSION BZ577315  
VERSION BZ577315.1 GI:27212376

KEYWORDS  
SOURCE

ORGANISM  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE  
1 (bases 1 to 891)

AUTHORS  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE  
Whole-Genome-Sequence variation among multiple isolates of

JOURNAL  
Pseudomonas aeruginosa library

COMMENT  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

## FEATURES

source  
1..891  
Location/Qualifiers  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="msh2\_5350"  
/clone\_lib="msh"  
/note="Environmental isolate. Whole genomic shotgun library."

## ORIGIN

Alignment Scores:  
Pred. No.: 6.16e-18 Length: 891  
Score: 272.00 Matches: 105  
Percent Similarity: 30.06% Conservative: 45  
Best Local Similarity: 21.04% Mismatches: 94  
Query Match: 10.52% Indels: 255  
DB: 8 Gaps: 11

US-10-009-823A-1 (1-502) x BZ577315 (1-891)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20  
Db 93 ATGTTATCGGCACTGTGGTTCAGCAACCGGTGTGCCGCCAGGACATGAACTGACC 152  
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyLysGlnGlnValVal 40  
Db 153 ACCATTTCACCAACACCTGGCCACAGTATCCACCACCGCTTCAAGCGGACCGCGGAG 212  
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60  
Db 213 TTCAGACCTGCTGTATCCAGATCCCGCGCCAGCCGCGCGCGAGTCGACCCAGGACAGC 272  
Qy 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79  
Db 273 GAGCTGCTTCGGGCGCTGCACTGGGTACCGGTGTGCGGTGTCGCGCACCCAGAGATC 332  
Qy 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyLys 94  
Db 333 TTCACCCCGGCGACCTGTCAGACACCGACCGACCGCGCTGACATGCGGTCAACGCGCGC 392



```

Db 801 AACAGCGTGTGGACCTGCCCATCAACGGCAACGGCTTCTCGTCACCAAGCAAAACACGG 860
QY 104 LysValHisTyrThrArgAlaGlyAsnPheA:gpHeThrGlnAspGlyPheLeuAsnAsp 123
Db 861 GCGATCAGGTACACCGCGCGCTACTTCAT-TACCGACAGCGGGATTCATTCCGCGAC 919
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLys--- 142
Db 920 CACAACCGTTACGNGCTGCAGGGCTTTGGCTCNGGCCGAACCGCCAGTTGCAGAACNGGT 979
QY 143 -----GluThr 144
Db 980 TGCTAACGACCTTCAAGGTGAGCGGCCCAATAGCGCGGAGCACTGGGCATCCAAAGG 1039
QY 145 LeuGluProIleGlnLeuAspPheAsn-----AspProThrValAlaLysSerPro 161
Db 1040 GTACAACCT---AAATTGAACCTTGAACCCCGACGCGGACCCCTTTATCCTT-----CCC 1090
QY 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
Db 1091 CCGCGTACTTAACCTGCTC-----TTGGTTGGGATT-----1123
QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
Db 1124 TTGAAGCGGTGCAATCCCAACCTTGGCTTTTATATAAAACACCGGTTCCCAAGTTACCC 1183
QY 202 ProfileSer 204
Db 1184 CCGGTTTCG 1192

RESULT 4
BH795484/c
LOCUS
DEFINITION
23H03L10003B Lawsonia library DNA linear GSS 21-OCT-2002
Genomic survey sequence.
ACCESSION
BH795484
VERSION
BH795484.1 GI:24182434
KEYWORDS
GSS.
SOURCE
Lawsonia intracellularis
ORGANISM
Lawsonia intracellularis
Bacteria; Proteobacteria; Deltaproteobacteria; Desulfotomobacteriales;
Desulfotomobacteriaceae; Lawsonia.
REFERENCE
1 (bases 1 to 742)
AUTHORS
Gebhart,C.J., Li,L., Zhang,Q., Guedes,R.M.C., Herron,L.L. and
Kapur,V.
TITLE
Genomic Sequence Survey of Lawsonia intracellularis, the causative
agent of proliferative enteropathy, identifies numerous genes of
relevance to diagnosis, virulence, and immunopathogenesis
JOURNAL
Unpublished (2002)
COMMENT
Contact: Zhang, Q
Department of Pathobiology
University of Minnesota
1971 Commonwealth Ave. St. Paul, MN 55108, USA
Tel: 612-625-9277
Fax: 612-625-5203
Email: qingzhang@umn.edu
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..742
/organism="Lawsonia intracellularis"
/mol_type="genomic DNA"
/db_xref="taxon:29546"
/clone_lib="Lawsonia library"
/note="Vector: pUC18; Site 1: SmaI"

ORIGIN
Alignment Scores:
Pred. No.: 1,22e-12 Length: 742
Score: 221.00 Matches: 62
Percent Similarity: 48.97% Conservative: 33
Best Local Similarity: 31.96% Mismatches: 87
Query Match: 8.55% Indels: 12
DS: 8 Gaps: 3

```

```

US-10-009-823A-1 (1-502) x BH795484 (1-742)
QY 316 GlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrp 335
Db 682 GGAAGTGCCTCACAACCGCTCATTAATTTTATCCAGGGGACTCACAGAAC-----632
QY 336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
Db 631 ACAGGTAACTCTTAATATGCTATTGAAGCAAGGGTTTTCACAGTA-----TTG 578
QY 356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
Db 577 GTTGGTGATCAATTAAATATGATACAGAGCAGGTGCTTTTAAACTTAATCAGGATGGACT 518
QY 376 ArgTyrSerGlnAspGlyTyrPro-----GlnGlyAspLeuVal 388
Db 517 CTGTACACGCCAATGATACCCCTCAACACAGAAATTTACAGTTCCTCCTGATCTAAA 458
QY 389 AspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
Db 457 TCTGTGCTGCTATTTCTGAAAGAGGTGCTATTACTGCTTTAGACGCAATGGACAAGAAATC 398
QY 409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGly 428
Db 397 GCTGCTGGAGAAATACCTTTATATAGCTTTGTTAATCCAGCGGGACTCGATCAAAAGGA 338
QY 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
Db 337 CGTAATCTCTATATTTCCCAAGACATCCGCTGAAGCGGTAGAGGTGCTCCCTGGTGA 278
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuThrSerAsnValAspMetSerArg 468
Db 277 GAAACTTAGGGACTATAGCTCAAGGATTCCTTGAATGCTCTAATGTAGAGGTGTTGAT 218
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db 217 GAAATGGTACAAATGATTTAGGACAAAGAGCATATGAATGAATTCNAAGCAATCCAA 158
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuLeuLysArg 502
Db 157 ACTTCAGACAAATGCTTCAAACTGCAGTCCAACTTAAACGT 116

RESULT 5
BZ563024/c
LOCUS
DEFINITION
pacs2-164_405.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_405, genomic survey sequence.
ACCESSION
BZ563024
VERSION
BZ563024.1 GI:27186093
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1105)
AUTHORS
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.
FEATURES
Location/Qualifiers
source
1..1105
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

```

```
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_405"
/library="pac2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."

ORIGIN
Alignment Scores:
Pred. No.: 1.26e-11 Length: 1105
Score: 214.00 Matches: 77
Percent Similarity: 47.03% Conservative: 34
Best Local Similarity: 32.63% Mismatches: 101
Query Match: 8.28% Indels: 24
DB: 7
Gaps: 7

US-10-009-823A-1 (1-502) x BZ563024 (1-1105)
QY 270 ThrMetThrPheSerSerAsnGlyGlu-----LeuLysAsnMetThrAlaPheThr 286
Db 1032 ACTTGCMAATTTCCACAACTGGTTCAAGGAGACCGCTGGATTCAGGAACTTTTACC 973
QY 287 ProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGly 306
Db 972 ATATAGACGACCAAGCGGGCGGCAATGTACCATGACAGGGAATCTCTCGCCCTGCAC 913
QY 307 LeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe 326
Db 912 TTGTTACCGTTTTCG-----GTGTCTCGGA-----CCAGAGATCTTAACCC 868
QY 327 Gly---IleLysSerGlnGlnAsnMetThrAlaGlyAlaProAlaSerAlaAlaIle 345
Db 867 GGAAGCTGAAGACGACGACGAGCGGTGG-----ACATTG 832
QY 346 GlyThrAspIleGlyLysLeuProSer-MetMet-----ProIleGlnThrSerSe 362
Db 831 CGGTAAAGGCGGCTTCTTCCAGCTCTGTGTGTCAGCGCACCTCTCTACACCGCG 772
QY 362 rGlyAsnSerThr-----AlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAs 380
Db 771 AGAAGCTCCACTTGAATCCGACGCGCAGATCGTCACTTCCAAAGCTCGCCCTGGAGCCAG 712
QY 380 pGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlyGly-LysLeuGlnG 400
Db 711 CGATCGTGTGCCATCGAGACCCAGACCTTCCCGTCGGCGGACGCGCTCTCTGTA 652
QY 400 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrS 420
Db 651 CCACACCGGCAACCGCCGCGCGAGTGTGCGCAACATCCAGACCGCGGACTTCATCA 592
QY 420 exGluAspGlyLeuArgArgGlyGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyG 440
Db 591 ACCGCGCGGCTGCGAGGCGCATCGGCAACACCTGTTCTCTGGAACCGGCTCCAGCGCG 532
QY 440 LysProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuG 460
Db 531 CGCCCGCAGGTGCGGTACGCGGCTTCAACGCGCGCTCGGCAAGTTCGCGCAACACCGCTG 472
QY 460 LuThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyP 480
Db 471 AAACTCTCAACGTCACCGTGTGCGAGAACTGGTGAACATCATCACCAACCGCGGCT 412
QY 480 hGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
Db 411 ACGAGATGAACCTCAAGGTGATCTTCCACCGCGCGACGAGATGTG 368

RESULT 6
LOCUS CF8911171
DEFINITION UI-CF-DUI-aas-f-05-18-UI.618 UI-CF-DUI Homo sapiens cDNA clone
CF8911171
ACCESSION UI-CF-DUI-aas-f-05-18-UI 3', mRNA sequence.
VERSION CF891117.1 GI:38150249

752 bp mRNA linear EST 03-NOV-2003
```

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 752)  
Ronald M. P., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/cf.html  
Seq primer: M13 FORWARD  
POLYA-No.

FEATURES  
source

Location/Qualifiers  
1..752  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI-aas-f-05-18-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_hosts="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-DUI"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-DUI is a normalized cDNA library containing the  
following tissue(s): Primary lung Epithelial cells The  
library was constructed according to Ronaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT73-pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dr)18 tail. The sequence tag for this  
library is GGCTGTAGGC.  
TAG\_SEQ=None found"

ORIGIN

Alignment Scores:  
Pred. No.: 2.34e-11 Length: 752  
Score: 209.00 Matches: 46  
Percent Similarity: 50.00% Conservative: 28  
Best Local Similarity: 31.08% Mismatches: 70  
Query Match: 8.08% Indels: 4  
DB: 7  
Gaps: 1

US-10-009-823A-1 (1-502) x CF891171 (1-752)

QY 357 ProfileGlnThr-----SerSerGlyAsnSerThrAlaArgAsnGlySerSer 372  
Db 8 CCAACACAGCGCGGTGCGCTCAACATGNGCGCAATCACCAGGTTCACCTCGGTACC 67  
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392  
Db 68 TACCGCAACCGCGCCACCGACCGGTTACGCCACCGGTCAATCACCAGGTGAAATC 127  
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412

```

Db 128 GAGCGGCGGGTGTGTTCGCCACTTTCAGCAACCCAGCAGCAAGGCCATCGGCCAG 187
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnHisTyr 432
Db 188 CTCCTCCCTGGCAGCTTCAACACGAGCAGGAGGCTTCGAGCAGCGGCGGACACCTGG 247
QY 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 248 AAGAAGAACCTTCGCGTGGCGCCAGCGGGTTTGTATACCCCGCAAGTCGGGACCTGGGT 307
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 308 TGAATCGTGGCACTCCCTGGAGAACTCCAACGCTCAACCTGACCAACGAGCTGTGGAC 367
QY 473 MetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThr 492
Db 368 CTGATCAAGGCCAGCAACTATCAGCGGAAGCAAGACCATCTCCACCCAAAGCACC 427
QY 493 MetLeuGlnLysAlaLeuGluLeu 500
Db 428 ATCATGCAACCATCATTCAGATG 451

RESULT 7
BZ577405
LOCUS BZ577405
DEFINITION msh2_5391.x1 msh Pseudomonas aeruginosa genomic clone msh2_5391,
 genomic survey sequence.
ACCESSION BZ577405
VERSION BZ577405.1 GI:27212466
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.
 Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) In press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.
FEATURES Location/Qualifiers
 source 1..895
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="M5H"
 /db_xref="taxon:287"
 /clone="msh2_5391"
 /notes="Environmental isolate. Whole genomic shotgun
 library."
ORIGIN
Alignment Scores:
Pred. No.: 3.47e-11 Length: 895
Score: 208.50 Matches: 48
Percent Similarity: 55.56% Conservative: 22
Best Local Similarity: 38.10% Mismatches: 49
Query Match: 8.06% Indels: 7
DB: Gaps: 3

```

US-10-009-823A-1 (1-502) x BZ577405 (1-895)

```

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHiserThrGlyLeuGly 20
 |||::: ::|||:::

```

```

Db 233 ATGTTATCGGCACTGTGGTTCAGCAAGACCGGTCTGTTCGCCCCAGGACATCAACCTGACC 292
QY 21 ThrValSerAsnLeuIleAlaAsnAlaAsnThrIleGlyTyrIlyGlnGlnValVal 40
Db 293 ACATTTTCCAAACCTTGGCCAACTGATCCACACCGGCTTCAAGCGGACCGCGCGGAG 352
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 353 TTCAGGACCTGCTGTATACCATGCTCCGCGCCAGCGGCGGCGGCTCGACCCAGGACAGC 412
QY 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79
Db 413 GAGCTCCCTTCGCGGCTGCAACCTGCGGTACCGGTGTGCGGTGCGGACCCAGAGATC 472
QY 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94
Db 473 TTCACCCCGGGAGCTCTGCAGACCCAGCAGCGCTGGACATGCGGTCAACGGGCGC 532
QY 95 GlyPhePheGlnValThrLeuGluAspLys---ValHisTyrThrArgAlaGlyAsnPhe 113
Db 533 GGCTTCTTCAGTCTCTGCTGCGCGGACGACCGGTGTCTACACCCGCGACGCGACTTT 592
QY 114 ArgPheThrGlnAspGly 119
Db 593 CACCTGAACCTCCGACGGG 610

RESULT 8
CB065273
LOCUS CB065273
DEFINITION EST644954 HOGA Medicago truncatula cDNA clone HOGA-1C13, mRNA
 sequence.
ACCESSION CB065273
VERSION CB065273.1 GI:27810851
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
REFERENCE 1 (bases 1 to 565)
 Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
 Utterback,T., Cho,J. and Fraser,C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished (2001)
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu
 TIGR sequence name: MTMAA19TK
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtg gat CC).
FEATURES Location/Qualifiers
 source 1..565
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="HOGA-1C13"
 /tissue types="3 day old seedling roots"
 /dev stage="24 hours after treatment in the dark at 26 C
 with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_hosts="XICLR"
 /clone_lib="HOGA"
 /notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging

```

extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in SOLR cells."

ORIGIN

Alignment Scores:  
Pred. No.: 2,75e-11 Length: 565  
Score: 206.50 Matches: 47  
Percent Similarity: 52.67% Conservative: 22  
Best Local Similarity: 35.88% Mismatches: 55  
Query Match: 7.99% Indels: 7  
DB: 6 Gaps: 3

US-10-009-823A-1 (1-502) x CB065273 (1-565)

QY 6 PheileGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsn 25  
Db 3 TGGGTCAGCAAGACCGGCTCGCGCTCATGACACCACTGACGACCATTTCCCAACAAC 62  
QY 26 IleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValPheGlnAspLeuPhe 45  
Db 63 CTGGCCACGCTCTCGACGACCGGTTTCAAGAAGGACCGCGCGAGTTTCAGGACCTGCTG 122  
QY 46 -----SerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn--- 60  
Db 123 TACCAGATCAAGCGCCACCGGTGCTCAGTCAGCCCAAGACGAGCGTGCACCGGT 182  
QY 61 ---GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAla 79  
Db 183 CTGACGTCGCTACTGCTGGCGCATCGTCGCGACCCCAAGAACTTCAACCGCGCGACG 242  
QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlySerThrGlySerGlnGlyPheGlnVal 99  
Db 243 CTGCAGACACCGACCGCGGTGGACATGGCCATCAACGGTCGGCGCTTTCAGATC 302  
QY 100 ThrLeuGluAsp---LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAsp 118  
Db 303 ATCAGCCCGGACGCGACCATCTCTATACCGGTGACGCGACCTTCACCTGACCTCCGAT 362  
QY 119 GlyPheLeuAsnAspProSerGlyPheThrLeu 129  
Db 363 GGTGAGTGTGTGACCGCCCAACGGTCTGCGCGCTG 395

RESULT 9  
LOCUS CO423013 898 bp mRNA linear EST 02-JUL-2004  
DEFINITION GGEZHT1021A02.9 HT1 Gallus gallus cDNA clone GGEZHT1021A02, mRNA sequence.

ACCESSION CO423013  
VERSION CO423013.1 GI:49639261  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 898)  
AUTHORS Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L.  
TITLE Discovery of new genes expressed in the chicken pituitary and hypothalamus

JOURNAL Unpublished (2004)  
COMMENT Contact: Clarissa S. Silva  
Laboratory of Animal Biotechnology, Dep. of Animal Production  
ESALQ - University of Sao Paulo  
Av. Pádua Dias, 11, Piracicaba, SP, 13418-900, Brazil  
Tel: 55 19 3429 4434  
Fax: 55 19 3429 4285  
Email: ccsilva@esalq.usp.br and llcouthin@esalq.usp.br  
PCR PRIMERS  
BACKWARD: T7.

FEATURES  
Location/Qualifiers  
1..898  
/organism="Gallus gallus"

/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="GGEZHT1021A02"  
/tissue\_type="pituitary and hypothalamus"  
/dev\_stage="21 days old"  
/lab\_host="DH10B"  
/clone\_lib="HT1"  
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This cDNA library was constructed with the SuperScript Plasmid System with Gateway Technology kit (Invitrogen), following a manufacturer's protocols. Plasmid DNA was purified using a modified alkaline lysis method. Sequencing reactions were conducted using the kit Big Dye Terminator Cycle Sequencing Ready Reaction (Applied Biosystems) according to the manufacturer's recommendations. Clones were sequenced by the 5' end with T7 primer. Sequencing reactions were analyzed on ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The quality and clustering of the ESTs were analyzed using the softwares Phred/Gap3. Only EST sequences with Phred quality greater than 20 and at least 150 bp were considered for clustering."

ORIGIN

Alignment Scores:  
Pred. No.: 2,47e-10 Length: 898  
Score: 200.50 Matches: 50  
Percent Similarity: 49.34% Conservative: 25  
Best Local Similarity: 32.89% Mismatches: 70  
Query Match: 7.75% Indels: 7  
DB: 7 Gaps: 3

US-10-009-823A-1 (1-502) x CO423013 (1-898)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23  
Db 431 TCTTTTAACTCGGCTTAGCGGCTCTATGCGGCCCAACAACAACCTGGACGTGACCGGC 490  
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValPheGlnAsp 43  
Db 491 AACAACTGCGCAACGTGCGGACCCCGGTTCAATCGTCCGCGCAATTCCTCGAT 550  
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63  
Db 551 ATCTACGCG-----GCGTCAAGCTGGCGCCGCCGACAGACAGCATCGGC 595  
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83  
Db 596 AACGGGTAAACCTGGCGGAGTGTCCAGCAGTTCACCCAGGCGTACGTCAACAACAGC 655  
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhe---GlnValThrLeuGlu 102  
Db 656 GCGCGCACGCTGAGCATGGCGATCCAGGCGCGCGGCTTCTTCGTGCAAGAGGCGAGCG 715  
QY 103 AspLysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122  
Db 716 GGTTCGCTGGAGTACACCCGTAAACGTCCTTCGCGCGCGACAAAGAGCGGTACATCACC 775  
QY 123 AspProSerGlyPheThr---LeuMetGlySerArgIleSerAsnAsnProAsnIleLys 141  
Db 776 AACAAACACCGCACCTCGCGCTGTCAAGGCTACGCGGCGNACGCGGACGCAACATCCAG 835  
QY 142 LysGluThrLeuGluProIleGlnLeuAspPheAsn 153  
Db 836 AAGGGCGGCTTGACGACCTTCAGCTCAACCTGTGCG 871

RESULT 10  
LOCUS BZ557649/c  
DEFINITION pacsl-60\_645.s1 pacsl-60 pseudomonas aeruginosa genomic clone  
ACCESSION BZ557649  
VERSION BZ557649.1 GI:27170747  
KEYWORDS GSS.

BZ557649 1190 bp DNA linear GSS 17-DEC-2002  
pacsl-60\_645.s1 pacsl-60 pseudomonas aeruginosa genomic clone  
BZ557649  
BZ557649.1 GI:27170747  
GSS.

```

SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
AUTHORS Pseudomonadaceae; Pseudomonas.
TITLE 1 (bases 1 to 1190)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.

FEATURES Location/Qualifiers
 1..1190
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone_lib="pacsl-60"
 /note="clinical isolate 1-60 Whole genomic shotgun
 library."

ORIGIN
Alignment Scores:
Pred. No.: 3.84e-10 Length: 1190
Score: 200.50 Matches: 58
Percent Similarity: 53.42% Conservative: 28
Best Local Similarity: 36.02% Mismatches: 49
Query Match: 7.75% Indels: 26
DB: 8 Gaps: 6

US-10-009-823A-1 (1-502) x BZ557649 (1-1190)
Qy 339 ProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu-----ProSerMetMet 356
Db 504 CCGCGAGCGGAGCTTCGACTGAACCTCCGAGCGGAGATCGTCACCTCCAGCGTTCG 445
Qy 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySer-SerSerThrArgAr 376
Db 444 CCT-----GGAGCCAGCGGATCGTGGTCCCAACGAGACCCAGAC 406
Qy 376 gTyrSer-----GlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSe 394
Db 405 GTTCACCGTCGGCCAGGACGGC-----ACCGTGTGGTGACAC 367
Qy 394 rGluglyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePr 414
Db 366 CACCGGCAACGCCCGAG-----CCGCGAGGTGATCGGC-----AACATCCA 328
Qy 414 oLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAl 434
Db 327 GACCGCGCACTTATACACCGCGCGCGCGGAAATCGATCGCTTGTCTCAACGCCCTCGGCACG 268
Qy 434 aThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLe 454
Db 267 AACCGGCTCCAGCGCGCGCGCGGAAATCGATCGCTTGTCTCAACGCCCTCGGCACG 208
Qy 454 uSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMetIl 474
Db 207 TGCCGAGAACACCTGGAAACCTCAACGTCACGTCGAGGTCGAGGAACTGGTGAACATGAT 148
Qy 474 eIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetIle 494
Db 147 CACCACCCAGCGCGCTACGAGATGAACCTCAAGGTCACTCCACCGCGCGGACCATGTT 88
Qy 494 u 494

Db 87 G 87

RESULT 11
CL660824/C
LOCUS
DEFINITION
Pseudomonas aeruginosa var. California Pristionchus pacificus genomic, genomic
survey sequence.
CL660824
CL660824.1 GI:50146561
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 849)
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

FEATURES Location/Qualifiers
 1..849
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN
Alignment Scores:
Pred. No.: 2.56e-10 Length: 849
Score: 200.00 Matches: 40
Percent Similarity: 65.17% Conservative: 18
Best Local Similarity: 44.94% Mismatches: 31
Query Match: 7.73% Indels: 0
DB: 9 Gaps: 0

US-10-009-823A-1 (1-502) x CL660824 (1-849)
Qy 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr 432
Db 844 ATTGACTGGCGGACTTGGCAACACGAGGCTGGCATCCGAGGCGACACGCTCG 785
Qy 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 784 TCTGCGACGCAATCTTCTGGCGTGGCGCTTTGGGACACGCCGGGAACTTTGGC 725
Qy 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 724 ACCCTGACCAACGGTGGCGCTGGAGCGCTCCAAAGTCAATCTCAAGAGACTGGTCAAT 665
Qy 473 MetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThr 492
Db 664 ATGATCGTTGCCAGCGTAACTATCATGCTTAAGCCGACCATCAAAACCCAGGCCAG 605
Qy 493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db 604 ATCTCAACACGCTGGTTAACTTACGC 578

```



Pred. No.: 3.95e-10 Length: 758  
 Score: 197.50 Matches: 65  
 Percent Similarity: 41.10% Conservative: 25  
 Best Local Similarity: 29.68% Mismatches: 84  
 Query Match: 7.64% Indels: 45  
 DB: 7 Gaps: 7

US-10-009-823A-1 (1-502) x C0421506 (1-758)

QY 299 GlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGly 318  
 DB 2 CAGGACAGCAATTGCCGTGGCGCTGCA-----CTGGGTACCGGT 43  
 QY 319 IleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAla 338  
 DB 44 GTGGCCATCGTTGGCACTCAGAGAAGCTTCAGCCCGGTAACTTGCAGCAAAACCGGCCAG 103  
 QY 339 ProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIle 358  
 DB 104 CCGTGGACTTGGGGTCAACGGTAAG-----GGTTCTTCAGATCCCTGCAGCCGAC 157  
 QY 359 GlnThr-----SerSerGlyAsnSerThr 366  
 DB 158 GGCACCACTACACCGCGCGGTACGTTCCACCTGGACGCAATGCCAGTCCGCTG 217  
 QY 367 AlaArgAsnGly-----SerSerSerThrArg 375  
 DB 218 ACCGCCACGGTTTCGGCTGGACCGCGGAGTTCGTGGGAAACAGCCGACACTTC 277  
 QY 376 ArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395  
 DB 278 ACCGTGGCAACGACGGC-----ACCGTTCGATCACTGTGGCC 316  
 QY 396 GlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu 415  
 DB 317 GGCAC-----CCGGCTCCCGAGTGATCGGC-----AACCTGCAAAACC 355  
 QY 416 AlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThr 435  
 DB 356 GCCGACTTCATCAACCGCGCGCTCGACGGGATGTTAACAACCTGTTCTCTGGAACC 415  
 QY 436 LeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer 455  
 DB 416 CGGTCCAGCGCGCGCAATCGGCAACCGCCCTTAACCGTTTGTGATCCACCCCTG 475  
 QY 456 ValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle 475  
 DB 476 CAAGACCCCTGGAAAGCTCCAGCTCAGACCGTTGAAGAGATGTTCAACATGATCACC 535  
 QY 476 IleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetLeu 494  
 DB 536 ACTCAGCGCGCTACGAGATGAACCTCAAGGTGATTTCCACCGCGCAGCAGATGCTT 592

RESULT 14  
 AJ747468 761 bp mRNA linear EST 07-JUL-2004  
 LOCUS  
 DEFINITION AJ747468 forward - stimulated minus unstimulated macrophage Sus  
 scrofa cDNA clone F\_C0001825b\_F11, mRNA sequence.

ACCESSION  
 AJ747468  
 VERSION  
 AJ747468.1 GI:49917717

KEYWORDS  
 EST.  
 SOURCE  
 Sus scrofa (pig)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE  
 1 (bases 1 to 761)  
 Hopwood, P.A., Zhang, F., Lowden, S., Talbot, R., Burt, D., Archibald, A.

AUTHORS  
 Development of a porcine cDNA microarray  
 Unpublished (2004)

TITLE  
 Contact: Hopwood PA  
 Dept. of Preclinical Veterinary Sciences  
 Royal School for Veterinary Studies

Summerhall, Edinburgh, EH9 1QH, UNITED KINGDOM  
 Sequencing was performed by ARK Genomics. This clone is available  
 from ARK- Genomics, Roslin Institute, Roslin, Midlothian EH25 9PS,  
 UK. See [www.ark-genomics.org](http://www.ark-genomics.org) or contact [info@arkgenomics.org](mailto:info@arkgenomics.org).

FEATURES  
 Location/Qualifiers  
 source  
 1..761

/organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /clone="F\_C0001825b\_F11"  
 /tissue\_type="lung"  
 /cell\_type="macrophage"  
 /clone\_lib="forward - stimulated minus unstimulated  
 macrophage"

ORIGIN

Alignment Scores:  
 Pred. No.: 3.97e-10 Length: 761  
 Score: 197.50 Matches: 47  
 Percent Similarity: 57.50% Conservative: 22  
 Best Local Similarity: 39.17% Mismatches: 44  
 Query Match: 7.64% Indels: 7  
 DB: 1 Gaps: 2

US-10-009-823A-1 (1-502) x AJ747468 (1-761)

QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeu-----GlnGly 400  
 DB 60 GCGGAATCGTTAAGTTCACCATCGTCTGATCGCTGGTGCAGGTAAACCAACAGGC 119  
 QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420  
 DB 120 CAGGCAGCTCCGTTTCAGTT-----GGGCAGCTCAATCTCACCACTTTATGAAT 170  
 QY 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440  
 DB 171 GACACCGGGCTGGAGAGCATTTGGGAAACCTCTACACCGAAACGCAATCTCTGGTGCA 230  
 QY 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460  
 DB 231 CGAAGCAAGACACCGCGGCTGACCGCGGAGCTGCTGTATCAAGGTATGTTGAA 290  
 QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPhe 480  
 DB 291 ACGTCTAAGCTCAACGTGGCGAAGAACTGCTCAATATGATTACAGTGCACGCGCTTAC 350  
 QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500  
 DB 351 GAATCAACAGTAAAGCGTGTCCACCAACCATCAGATGCTGCAAAACTGACCGCAACTC 410

RESULT 15  
 C0423511 766 bp mRNA linear EST 02-JUL-2004  
 LOCUS  
 DEFINITION GGEZHT1017H04.g HT1 Gallus gallus cDNA clone GGEZHT1017H04, mRNA  
 sequence.

ACCESSION  
 C0423511  
 VERSION  
 C0423511.1 GI:49639759

KEYWORDS  
 EST.  
 SOURCE  
 Gallus gallus (chicken)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

REFERENCE  
 1 (bases 1 to 766)  
 Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L.

AUTHORS  
 Discovery of new genes expressed in the chicken pituitary and  
 hypothalamus

JOURNAL

Unpublished (2004)

COMMENT  
 Contact: Clarissa S. Silva  
 Laboratory of Animal Biotechnology, Dep. of Animal Production  
 ESAUQ - University of Sao Paulo  
 Av. Padua Dias, 11, Piracicaba, SP, 13418-900, Brazil  
 Tel: 55 19 3429 4434



Db 658 GCGGCGACGCTGGACATCGCATCAGCGCGCGCTCTTCGTGCGAGAGGGGCGAGC 717  
 Qy 103 AsplysvalHistyThrArgAlaGlyAsnPhaArgPheThrGlnAspGlyPheLeuAsn 122  
 Db 718 GGTTCGCTGGAGTACACCGGTAACGGTTCCTTCGCGCGGCGGTAAGACGGGTACATCACC 777  
 Qy 123 AspProSerGlyPheThr---LeuMetGlySerArgIleSerAsnProAsnIleLys 141  
 Db 778 AACACACGGGACCTCGCGCTGACAGGCTACGCGGCGGACCGGACGCGACATCCAG 837  
 Qy 142 LysGluThrLeuGluProIleGlnLeuAsp 151  
 Db 838 AAGGGCGCTTGACCGGACTTCAGCTCAAC 867

RESULT 17  
 CB854175 650 bp mRNA linear EST 22-APR-2003  
 LOCUS  
 DEFINITION UI-CF-DUI-aas-f-05-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone  
 CB854175  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 650)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccrayuiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com) or from Open Biosystems  
 Genetics (www.openbiosystems.com).  
 Seq primer: M13 FORWARD  
 POLYA=No.

FEATURES  
 source  
 1..650  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-DUI-aas-f-05-0-UI"  
 /tissue\_type="Primary Lung Epithelial Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-DUI"  
 /notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-DUI is a normalized cDNA library containing the  
 following tissue(s): Primary Lung Epithelial Cells The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this

library is GGCTGTAGG.  
 TAG\_SEQ=None found"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.08e-08 Length: 650  
 Score: 183.00 Matches: 46  
 Percent Similarity: 49.33% Conservative: 28  
 Best Local Similarity: 30.67% Mismatches: 70  
 Indels: 6  
 Gaps: 1  
 DB:

US-10-009-823A-1 (1-502) x CB854175 (1-650)

Qy 357 ProIleGlnThr-----SerSerGlyAsnSerThrAlaArgAsnGlySerSer 372  
 Db 9 CCACACAGACGGCGCGTGCCTCAACATGNGCGGAACATCACCAGTTCAACTCGGTACC 68  
 Qy 373 SerThrArgArgTySerGlnAspGlyTyProGlnGlyAspLeuValAspValThrIle 392  
 Db 69 TACCGCAACCGCGCCACCACCGAGCGTTACGCCCGGTCAAAATCACCGGCGCTGAAAATC 128  
 Qy 393 ThrSerGluGlyLysLeuGlnGlyLysTySerAsnSerGlnValValAspPheTyAsn 412  
 Db 129 GAGCGACGCGGTGTGTTCGCCCACTTTCAGCAACACGACGACGAGCGGCATCGGCCAG 188  
 Qy 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTy 432  
 Db 189 CTCCTCCCTGGCCAGCTTCAACACGACGAGCGGCTCGACGCGGCGGCGGACCATCTGG 248  
 Qy 433 -SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTy 452  
 Db 249 GAAAGAAACCTTCGCGTGGCGCGCGGTTTGATACCCCGCAAGTCGCGGACCCCTGGG 308  
 Qy 452 YLysLeuSerValAsnGlnLeu-GluThrSerAsnValAspMetSerArgGluMetVala 472  
 Db 309 TTCGATCGTGGCCCACTCCCTNGGAGAACTCCAACGTCAACCTGACCAAGAGTGTGG 368  
 Qy 472 snMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspT 492  
 Db 369 ACCTGATCAGGCGCCAGACCAACTATCAGCGCAAGCGCAAGACCATCTCCACCCAAAGCA 428  
 Qy 492 hrMetLeuGlnLysAlaLeuGluLeu 500  
 Db 429 CCATCATGCAGACCATCATTCAGATG 454

## RESULT 18

CN582719 661 bp mRNA linear EST 04-MAY-2004  
 LOCUS  
 DEFINITION USDA-FP\_125782 Acyrthosiphon pisum, Pea Aphid Acyrthosiphon pisum  
 CN582719 cDNA clone WHAP021\_G07, mRNA sequence.

ACCESSION CN582719

VERSION CN582719.1 GI:46994442

KEYWORDS

SOURCE

ORGANISM

Acyrthosiphon pisum (pea aphid)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;

Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

1 (bases 1 to 661)

Hunter,W.B., Dang,P.M., Stern,D. and Lapointe,S.L.

Gene expression in Acyrthosiphon pisum (Hemiptera: Aphididae)

Unpublished (2004)

Contact: Wayne Hunter; Phat Dang

USDA-ARS

US Horticultural Research Lab, Ft. Pierce, FL

2001 South Rock Rd., Ft. Pierce, FL 34945, USA

Tel: (772) 462-5898, (772) 462-5940

Fax: (772) 462-5986

Email: Whunter@ushrl.ars.usda.gov.

Location/Qualifiers

1..661

/organism="Acyrthosiphon pisum"

```

/mol_type="mRNA"
/db_xref="taxon:7029"
/clone="WHA021_G07"
/sex="Mixed population"
/tissue_type="whole body, nymphs and adults"
/dev_stage="All"
/lab_host="X11-Blue"
/clone_lib="Acyrthosiphon pisum, Pea Aphid"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; Acyrthosiphon pisum colonies. Library by Srin Kambampati; Peter Dearden; David Stern, Department of Ecology and Evolutionary Biology, Princeton University, NJ. (609) 258-0759. A high quality EST with at least 100 contiguous bases at Trace Tuner score of 20 or better. Sequencing of clones by Dr. PM Dang, US. Horticultural Research Lab, Ft. Pierce, FL. (772) 462- 5940."

ORIGIN
Alignment Scores:
Pred. No.: 1.6e-08 Length: 661
Score: 181.50 Matches: 58
Percent Similarity: 43.35% Conservative: 30
Best Local Similarity: 28.57% Mismatches: 88
Query Match: 7.02% Indels: 27
DB: 7 Gaps: 6

US-10-009-823A-1 (1-502) x CNS82719 (1-661)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db TCAATATGATAGCTATAGTCTAGCTTACTAGCAATATGATACATGGAATATATCC 540
QY 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrGlyGlnGlnValValPheGlnAsp 43
Db AATAATATTCCTAACGCATCACTATAGGATATAAATCTCGTAAACCTCTCTTTTGTAT 480
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db ATGTTTCTCATTCATTTTATCAATATCAATATCACTAAAT-----GGA 441
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db TACGAGTGGGTATTTCAAGTATATATACAAACTTTTAATATGTCATGTTAGTGAACCT 381
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPheGln---ValThrLeuGlu 102
Db GGACGAGATTGGATTATAGGAATATATAAAGACGGCTTTTTCGTCCTTGAGACATCAA 321
QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
Db GGTCTGCTATATATACAGAGATGGCAATTTCTCTCGATAAAGATCAAAATATATATC 261
QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136
Db AATATTCAGGTATGATATCACTGGACTTAATATACATCTTGTTCAAAAAGTATTTTAA 201
QY 137 AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156
Db -----AGATCCAACTTTAGAACCTTATTAATTA-----AAAAATCTAAT 159
QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly--- 175
Db ATTTTAAAAACAAACCTTACTTCTGAAATCATGTATTAAGGCTTTTGTGAATCGTAATCT 99
QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189
Db GAATCAAAACAGCTGGTGAATATCTTGACAAACAACTATCTATAACCAAGAGATTACATG 39
QY 190 AlaLeuLeu 192
Db 38 ACTTATATT 30

```

RESULT 19

```

BP562788/c
LOCUS BP562788 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-76-G17 5',
DEFINITION mRNA sequence.
ACCESSION BP562788
VERSION BP562788.1 GI:48978554
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 566)
AUTHORS Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Ono, Y., Muramatsu, M.,
Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.
FUNCTIONAL annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
TITLE Science 296 (5565), 141-145 (2002)
JOURNAL Science
MEDLINE 21932300
PUBMED 11910074
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-3060
Email: msekior@c.riken.go.jp
Please visit our web site (http://pfweb.gsc.riken.go.jp/) for
further details.
FEATURES
Location/Qualifiers
source 1..566
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL14-76-G17"
/tissue_type="root"
/lab_host="DH10B"
/clone_lib="RAFL14"
/notes="Site_1: BamHI; Site_2: SalI"

ORIGIN
Alignment Scores:
Pred. No.: 2.95e-08 Length: 566
Score: 178.00 Matches: 36
Percent Similarity: 62.65% Conservative: 16
Best Local Similarity: 43.37% Mismatches: 31
Query Match: 6.88% Indels: 0
DB: 5 Gaps: 0

US-10-009-823A-1 (1-502) x BP562788 (1-566)
QY 418 PheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAsp 437
Db TTTTATGATGACACCGGCTGGAGAGCATTTGGGAAACCTCTACACGAAACGCAATCC 451
QY 438 SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLeuSerValAsn 457
Db TGTGTGTGCACCAAGCAACGACGCGGCTGTAACGCGCGGAGCTGCTGTATCAAGGG 391
QY 458 GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGln 477
Db TATGTTGAAACCTCTTAACGTCACGTCACGTCGCGGAGAACTGGTCAATATGATTCAGTGC 331
QY 478 ArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetIleGlnLysAla 497
Db CGGCTTACGAATCAACAGTAAAGCGGTGTCCACCACCGATCAGATGCTGCAAAACTG 271
QY 498 LeuGluLeu 500
Db 270 ACGCAACTC 262

```

RESULT 20

```

CD119056
LOCUS ME1-0052U-V142-G10-U-B 436 bp mRNA linear EST 14-SEP-2003
DEFINITION ME1-0052U-V142-G10-B, mRNA sequence.
ACCESSION CD119056
VERSION CD119056.1 GI:34657168
KEYWORDS EST;
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
REFERENCE 1 (bases 1 to 436)
AUTHORS Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
 Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
 Kitajima, J.F., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
 Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
 Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
 Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
 Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
 Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
 Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
 Transcriptional analysis of the acoelomate human parasite Schistosoma
 mansoni
 Nat. Genet. 35 (2), 148-157 (2003)
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE 22879926
PUBMED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
 Departamento de Bioquímica
 Instituto de Química - Universidade de São Paulo
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
 Brasil
 Tel: +55-11-3091-2173
 Fax: +55-11-3091-2186
 Email: verjoe@iq.usp.br
 This sequence was derived from the FAPESP Schistosoma mansoni EST
 Genome Project. All sequences in the project were assembled and
 annotated. This entry and all the assembled sequences can be seen
 in the following URL http://bioinfo.iq.usp.br/schisto/
 Plate: ME1-0052U-V142 row: 10 column: G.
FEATURES Location/Qualifiers
 source 1..436
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /db_xref="taxon:6183"
 /clone="ME1-0052U-V142-G10-B"
 /sex="mixed pool"
 /dev_stage="egg"
 /lab_host="Mus musculus"
 /clone_lib="ME1-0052"
 /note="vector: pGEM T-easy"
ORIGIN
Alignment Scores:
Pred. No.: 4.08e-08 Length: 436
Score: 175.00 Matches: 33
Percent Similarity: 62.92% Conservative: 23
Best Local Similarity: 37.08% Mismatches: 33
Query Match: 6.77% Indels: 0
DB: Gaps: 0

US-10-009-823A-1 (1-502) x CD119056 (1-436)

QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnHisTyr 432
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2 GTTCGTCGTGACACCTTTTGATATGCCAGGCTGCAGCCCAATGGCGAAACAACTGG 61
QY 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 GCATCTACTCCGCTTCGGGTGAGGCAAGCCCTGGCTCCCGGAGAGGAGGATTITGC 121
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 122 AAGTTGCAATCGCTGTCATTTGAAGAATCCACCGGATACCACCGCAGAGCTGATCAAC 181

```

```

Db 2 GTTCGTGTACACCTTGTATGTAATGAGCCAGCGCTCGAGCCCAATGCGGAAACAACTGG 61
Qy 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 62 GCATCTACTCGCGCTCGGTGAGCCCAAGCTGCTTCGCGGAGAGAGCGAGTTCGCTTGGC 121
Qy 453 LysLeuSerValAnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 122 AAGTTGCAATCGTCTGCGATTTGAAGATCCCAACGCCGATACACGCGAGAGCTGATCAAC 181
Qy 473 MetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 182 ATGATGTCGCGAGCGCTTCTATCAGCCCAAGCGCAGACCATCAAGACGCGAGTGGC 241
Qy 493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db 242 ATCTGCAACCCCTGCTCAACCTCGT 268

RESULT 22
LOCUS AI057966/c
DEFINITION SWOV13CAN21B03SK Onchocerca volvulus infective larva cDNA
(SW94WL-OVL3) Onchocerca volvulus cDNA clone SWOV13CAN21B03 5',
mRNA sequence.
ACCESSION AI057966
VERSION AI057966.1 GI:3331832
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE Williams,S.A., Lu,W., Lizotte-Waniewski,M. and Laney,S.J.
TITLE Genes expressed in infective third stage larvae of Onchocerca
JOURNAL Unpublished (1995)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK
Location/Qualifiers
1..532
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="SWOV13CAN21B03"
/lab_host="XL1-Blue MRP"
/clone_lib="Onchocerca volvulus infective larva cDNA
(SW94WL-OVL3)"
/note="vector: lambda UniZap XR; Site 1: EcoR I; Site 2:
Xho I; Cutaneous filarial nematode parasite of humans.
mRNA was prepared from third stage infective larvae of
Onchocerca volvulus isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNase I. The library had 1.8 x 10^8 independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is
available from Dr. S.A. Williams, email genome@smith.edu."

ORIGIN
Alignment Scores:
Pred. No.: 1.03e-07 Length: 532
Score: 172.50 Matches: 64
Percent Similarity: 39.57% Conservative: 29
Best Local Similarity: 27.23% Mismatches: 73

```

```

Query Match: 6.67% Indels: 70
DB: 1 Gaps: 8
US-10-009-823A-1 (1-502) x AI057966 (1-532)
Qy 216 ValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSer 235
Db 516 GTTTCGACAGTCAGGTATGTCATGACATGAGCGTCTACTTT-----GTG 469
Qy 236 SerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySer 255
Db 468 AAGACCGCGGATATACTACCTGCGAG-----GTCTACACCCGAGATAGCAGTCCAAAC 415
Qy 256 AlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSer 275
Db 414 AGCATTGCGAAGACA-----GCGACACACANTGGAAATTACGCT 376
Qy 276 AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeu 295
Db 375 AATGCGCACATTA-----GTGGATGGTGGCGATGGCGAATAATATCCCAACCGCGCAATT 322
Qy 296 AsnAlaThrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheVal 315
Db 321 AACGTCACAGACCGGCC-----ACGTTTAGTCTGAGCTTCCTC 283
Qy 316 GlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyr 335
Db 282 AACTCC-----ATGCAGCAAAAT-----265
Qy 336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
Db 265 -----265
Qy 356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
Db 264 -----ACCGCGCTAAACAATATTGTGGCAAC-----238
Qy 376 ArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
Db 237 -----ACCAGAACGAGNAC-AAACCGCGGATCTGGTGATTCATCAATCAATGATGAC 185
Qy 396 GlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu 415
Db 184 GGTACGGTTTGTGCGCACTATTCCACAGAACAAACCAACTGTCTGGGCGAGATTGACTG 125
Qy 416 AlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThr 435
Db 124 CGGAACNTTGCACACAGAGGTCTGGCATCCGAGGGGAGACAACTCTGCTGTGGACG 65
Qy 436 LeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db 64 CAATCTTCTGCGTGGCGCTGTTGGGACAGCGCGGTGCACGTCA 20

RESULT 23
LOCUS CD444654
DEFINITION ELO10N0442D08.b EndospERM_4 Zea mays cDNA, mRNA sequence.
ACCESSION CD444654
VERSION CD444654.1 GI:31360297
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 715)
Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,B. and
Messing,J.
REFERENCE Sequencing of the maize endospERM ESTs
AUTHORS Unpublished (2002)
JOURNAL Contact: Lai, Jinsheng
COMMENT Dr. Joachim Messing's lab
Waksman Institute, Rutgers University

```



```
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
source
1. .1158
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_658"
/clone_lib="pacs1-60"
/notes="clinical isolate 1-60 Whole genomic shotgun
library."
ORIGIN
Alignment Scores: 5.14e-06 Length: 1158
Pred. No.: 83
Score: 161.50
Percent Similarity: 39.88%
Conservative: 49
Best Local Similarity: 25.08%
Indels: 92
Query Match: 6.25%
Gaps: 19
DB: 8
US-10-009-823A-1 (1-502) x BZ557676 (1-1158)
QY 200 ThrPro-----ProlleSerThrSerAsnTySerTyAlaGlnProMetArgValtyr 217
Db 69 ACGCCGTTTCGATCGTCGACGCGCTACCTACAACTGCTCTTCTGCTGGCGATCTAT 128
QY 218 AspGlnGlnGlyAsnSerHisAspIleThrValTyPhe-----AspGlyAlaPro 234
Db 129 GACAGCCAGGCGCACTCCACACCATGACGAGTTCATCAAGAACGAGCCGGATCCG 188
QY 235 SerSerThr-----GlySerTySerPheGluTyLeuVal---AlaMet 248
Db 189 AATCGACCCCGCGATTCGAGAACAGCTGGACCATGAAGTCTGATCGACGGGTC 248
QY 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer 268
Db 249 AATCGCTCGACCCGTCGACACAGACGCG-----ATGAGC 284
QY 269 GlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThr 288
Db 285 TTCACGTCACCTTCGACCGCGCGCGAGATGACCTCGGTACGGGT-----CCGGAC 338
QY 289 GlySer-----AlaThrTySerAspLeuAsnAlaTrpGln 299
Db 339 GGCAGCACCGCGCGCGGCTTCAGCATCGACGCGACCAACAGTGATCCAGTTTCAGC 398
QY 300 ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLe 319
Db 399 CCGGCC-----ACCGGAATCCCGCGCGACTCC-----GGCACCAGGCTGG 437
QY 320 GlnProLeuThrLeuAspPheGlyLeuLysSerGlnGlnAsnMetTrpAlaGlyAlaPro 339
Db 438 ATTCGCGCGGCTCGGAC---GGCAAGACCCCGCGCGACTACGCTGGAATGGCGGAC 494
QY 340 AlaSerAlaAlaAlaIleGlyThrAspIleGlyLeuProSerMetMetProIleGln 359
Db 495 GGTGCGCGCGCGGATCTCTCTTCGACATGCGCAAG-----530
```



```

ACCESSION CN822868
VERSION CN822868.1 GI:47950937
KEYWORDS EST.
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Caprinae; Ovis. 1 (bases 1 to 747)
AUTHORS Gosner, A. and Hopkins, J.
TITLE Ovine spleen/brain cDNA library
JOURNAL Unpublished (2004)
COMMENT Contact: J Hopkins
 Veterinary Biomedical Sciences
 University of Edinburgh
 Summerhall Square, Edinburgh, EH9 1QH.
 Email: j.hopkins@ed.ac.uk
 Plate: 02 row: E column: 20
 Seq primer: M13reverse
 High quality sequence start: 7
 High quality sequence stop: 549.
 Location/Qualifiers
FEATURES 1..747
 /organism="Ovis aries"
 /mol_type="mRNA"
 /db_xref="taxon:9940"
 /clone="Oa脾bn.02E20"
 /clone_lib="Sheep spleen/brain pSport1 library"
ORIGIN
Alignment Scores:
Pred. No.: 1..628-05 Length: 747
Score: 154.00 Matches: 39
Percent Similarity: 46.43% Conservative: 13
Best Local Similarity: 34.82% Mismatches: 54
Query Match: 5.96% Indels: 6
DB: 7 Gaps: 2

US-10-009-823A-1 (1-502) x CN822868 (1-747)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
DB 426 TCTTTTAACTGCGGCTTATGCGGCTCTATCGGCCCAACAACTGCGTGCACGCGC 485
QY 24 AenAenIleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValPheGlnAsp 43
DB 486 AACACATGCGCAAGTCGCGACCAACCGCTTCAGTCGTCCTCGGATTCCTCGAT 545
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
DB 546 ATCTACGGCG-----CGGTCCAAGCTGGCGCCCGCCAGAACAGCATCGGC 590
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
DB 591 AACGCGTAAACCTGGCGGAGTGTCCAGCAGTTCACCGAGGTGAGTCAACACAGC 650
QY 84 AenSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhePhePhePhePhePhePhe 102
DB 651 GCGCGCAGCGTGGACATGCGCATCGCGGCGCGGCTTCGTCGACAGAGCGGCGAC 710
QY 103 AspLysValHisThrThrArgAlaGlyAsnPheArg 114
DB 711 GGTTCGCTGGAGTACACCGTAAACGTCGCTTCCTCCG 746

RESULT 29
AK044947
LOCUS
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
 [Rattus norvegicus], full insert sequence.
ACCESSION AK044947
VERSION AK044947.1 GI:26336942
KEYWORDS HTC; CAP trapper.

```

```

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL 20499374
MEDLINE 11042159
PUBMED
REFERENCE Carninci, P. and Hayashizaki, Y.
AUTHORS Carninci, P., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Izawa, M., Ohara, E., Matsuura, S., Kawai, J.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multipillar sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL 20530913
MEDLINE 11076861
PUBMED
REFERENCE The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
JOURNAL 6 (bases 1 to 3057)
MEDLINE
PUBMED
REFERENCE Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
JOURNAL
MEDLINE
PUBMED
COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. Tomohiro Kono (Department of Animal
 Science, Tokyo University of Agriculture, 1737 Funako Atsugi City,
 Kanagawa Prefecture, Japan) whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.

```

```

URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
 source
 1..3057
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:BI30015D15"
 /db_xref="taxon:10090"
 /clone="BI30015D15"
 /tissue_type="parthenogenote"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="9.5 days embryo"
 1..2002
 /note="unnamed protein product; nucleoporin homolog
 [Rattus norvegicus] [PIR|A44345, evidence: PASTY, 92.8%ID,
 45.4%length, match=1995]
 putative"
 /codon_start=2
 /protein_id="BAC32155.1"
 /db_xref="GI:26335943"
 /translation="CNKAEDNRVCSTSEKPLVLSASSSPAPVLSGGCLGLDKFK
 KPEGSWDCEVLQNKADSAKCIACESAKPGTKSEFGFGTSSSLNPAPSAFKGIPF
 SSGSLQTLSTGNFKFGDGGFLGTSSSDSGSTNTMTNFKFSGKPEELPKSPAGPSPA
 KPEEVKNDKNDTFQSGSLTNPASSAPQFGVTLTGQOKEELPKSPAGPSPA
 AGVNNPNAADITATSENKSGENFGFLDTKSVSVTPETVTKTEAKKEDAPATKGF
 FGKVGSSSLPSSMFLVGRTEKQOEPVTSLSVFGKADSEEPKCPQVSPFNSSQGT
 KDESKPFPFSVAKPGSEKSEQLAKATFAPGNQNTITDQGAAPVPSFLNSSSSS
 SAPATSSGGPIFGSTSSNPVAAFFVQGNPVSFAFGNAESSTQSLLFPQES
 EPATSSAPAAFPVFTGASSNSVSSGFTGATTTSSSGSFVFGTGHSAFASAP
 AFGANQPTFCQSCAGQNPFPFGSISSTALFASAGQVPVPIFGTVSSSQPVF
 CQOPQSQAFSGGTANAGSVFQSGTNNFNFTNPNPSGVTFGASPTPAASQPSGS
 GVFSQSQSPASFTVSGSKNMFSSSGTSVSGRIKIKTAVRRK"
 3039..3044
 /note="putative"
 3057
 /note="putative"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00579 Length: 3057
 Score: 139.00 Matches: 122
 Percent Similarity: 31.66% Conservative: 48
 Best Local Similarity: 22.72% Mismatches: 217
 Query Match: 5.38% Indels: 150
 DB: 3 Gaps: 23
 polyA_signal
 polyA_site
 US-10-009-823A-1 (1-502) x AK044947 (1-3057)
 QY 22 ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe 41
 DB 494 GTTAAACACGACACAAAGATGATAAT-----TTT 523
 QY 42 GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln 61
 DB 524 CAG-----TTGGATCTCTTCGGCTTAACCAACCCAGCTTTCAGCTCCGTTTCAG 577
 QY 62 AlaGlyMetGlyAla-----GlnValGlySer 70
 DB 578 TTTGGGTATCTACTCTTGGCAACAAAGAAAGAGAGAGCTGCCTAAATCTTCACCT 637
 QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
 DB 638 GCAGGCTTTAGCTTGTCTGCGAGGTGTTAATAACCCCTTAATGCTCTATTGACACCACA 697
 QY 91 IleGlyGlyLys-----GlyPheGlnValThrLeuGluAspLys---ValHis 106
 DB 698 GCGACCTCAGAGAACAGACGCGCTTCACTTTGGAAACCCCTAGACACAAAGAGTGTCTCA 757
 QY 107 TyrThr-----Argala 110
 DB 758 GTGACTCTTTTCATACAGACACAGAGCAAGAAAGAAAGAGACGACCTGCTACCAAG 817
 QY 111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet 130
 DB 818 GGTGGGTTCACATTGGTAAAGTGGGCTCTCTCTTCTTCCATCTTCCATGTTGT 877
 QY 131 GlySerArg----- 133
 DB 878 TTGGGAAGGACAGAGAGAGAAACAGCAAGAACCTGTTACTTCTCTGTTGTTGGG 937
 QY 134 -----IleSerAsnAsnPro 138
 DB 938 AAGAAAGCCGACAGTCAGGAGCAAGAGTCTCAGCCAGTATTTCTTCTTGGGAATTCAGAG 997
 QY 139 AsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAla 158
 DB 998 CAAACCAAGATGAGAGCTCCCAAGCG--ACATTGAGCTTC-----AGTGTGGCA 1045
 QY 159 LysSerProAlaLysThrSerThrAlaLeu---AsnAlaValValAsnLeuGlyAspSer 177
 DB 1046 AAACCACTCTGGGAAGAGTCTCAGCAGCTAGCCAGGCTATCTTTGGCATTTTGGAAATCAA 1105
 QY 178 ThrAspLysThr-----GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyr 195
 DB 1106 ACCAATACAACTGATCAAGGTGCAGCAAGCCAGTTTTTAGCTTCTTGAACAGCAGT 1165
 QY 196 LysGlyAsnGlyThrProProIle-----SerThr 205
 DB 1166 TCCTCTAGTCAAGTGCACCCAGCTTCACTCTAGTGTGGCATATTTGGTAGTTCCACC 1225
 QY 206 SerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225
 DB 1226 TCTTCTCCAAACCCAGCTGTGCAGCCTTTGTGTGTGACAGCCAGCAATCTCTCAGC 1285
 QY 226 IleThrValTyrPheAspGlyAlaProSerSerThrGly----- 238
 DB 1286 AGCTCTGCTTTCGGTAAACGCTGCTGAATCCAGTATCATCTCAGTCTTTGCTATTTCTCAA 1345
 QY 239 -----SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
 DB 1346 GAGAGCGAGCCAGCCAGCTCCAGCAGCTCCCGCAGCTCTCTCTTGTATTGTTGGG 1405
 QY 255 SerAlaAlaSerGlyThrAsp---SerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 1406 ACAGAGCGAGCAGCAGTAAATAGTGTCTCTCTGGGTCTCACTTTGGGGTCTACAAACGCTCA 1465
 QY 274 SerSerAsnGlyGluLeu-----LysAsnMetThr 283
 DB 1466 AGCTCTTCAGGATCTCTCTTGTATTGGCACTCGACATTCAGCACCATTCGCGAGTCCA 1525
 QY 284 AlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGln 299
 DB 1526 GCATTGGTGTAAACAGACTCCAACTTTGGCAAAAGTCAA-----GGTGCAGCCAG 1579
 QY 300 ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLe 319
 DB 1580 CCTAACCCGCAAGCTTGGCTCTATCTGCTTCAACAGCATTAATTTCTGCTGTTCT 1639
 QY 320 GlnProLeuThrLeuAsp---PheGlyLe-----LysSerGlnGlnAsnMetTyrAla 336
 DB 1640 CAGCTGTACCACCCACTATATTGGGACAGTGTCAAGCAGCAGCCAACTCTCTGTTT 1699
 QY 337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
 DB 1700 GGACAGCAGCCCTAGCTAGTCTGCAATTTGGCTCTGGG----- 1735
 QY 357 ProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArg 376
 DB 1736 -----ACAGTAAAGCAGTTCGTTTCCATTTGGTACGACCTACGAATTC 1786
 QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGly 396
 DB 1787 AACTTCAAAACATAATATCCATCAGGAGTGTTCACATTTGCTGCAAGTCCAGCAGCCT 1846
 QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416

```

```

Db 1847 GCAGCTCAGCCAGCCCTCAGGTCAGGGGTCTTTTCATTACGCGAGTCTCCA---GCA 1903
Qy 417 ArgpThrSerGluAspGlyLeuArgGluGluGlyAsnAsnHisTyrSerAlaThrLeu 436
Db 1904 TCATTTTACA-----GTGGGGTCAAAATGGGAAATATGTTCTCTCTTC--- 1948
Qy 437 AspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 1949 -----GGAACTTCAGTTCTTCTGGTCGC 1969

RESULT 30
LOCUS AZ577776
DEFINITION 14f11 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
ACCESSION AZ577776
VERSION 1
KEYWORDS 14f11, genomic survey sequence.
SOURCE Rhizobium sp. NGR234
ORGANISM Rhizobium sp. NGR234
REFERENCE 1 (bases 1 to 374)
AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
TITLE Genetic snapshots of the Rhizobium species NGR234 genome
JOURNAL Genome Biol. 1 (6), RESEARCH0014 (2000)
MEDLINE 21114532
PUBMED 11178268
COMMENT Contact: Virginie Viprey
 Laboratoire de Biologie Moleculaire des Plantes Superieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.viprey@bsrc.ac.uk
 Class: shotgun.
FEATURES Location/Qualifiers
 source 1..374
 /organism="Rhizobium sp. NGR234"
 /mol_type="genomic DNA"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="14f11"
 /clone_lib="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="Vector: ML3; derivative strain of NGR234 cured of
 pNGR234a"
ORIGIN
Alignment Scores:
Pred. No.: 0.000275 Length: 374
Score: 138.00 Matches: 28
Percent Similarity: 56.76% Conservative: 14
Best Local Similarity: 37.84% Mismatches: 32
Query Match: 5.34% Indels: 0
DB: 8 Gaps: 0

US-10-009-823A-1 (1-502) x AZ577776 (1-374)
Qy 420 SerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db 5 AACGAGCGCGGTGGAGCCGCGAGCGGCGAAACCTCTCAAGCAGACAGCGCGATCCGGC 64
Qy 440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db 65 GAACCGATCGTCGGCAGCGCGCGCGATCCCGGCTCCGCGCAGATCAAGCAGGGCTATCTG 124
Qy 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGly 479
Db 125 GAATCTTCAAGTCGATCGGTCAAGGAATACCCGACCTGATCTCGCGCCAGCGCGCC 184
Qy 480 PheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493

```

```

Db 185 TACGAATGAATCAAGTCTCATCCAGCGCGGACGAGATG 226
Qy 190 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 169 ACTGG---AAGGGTGACTCCAGCGGCTTCAGTAGCTACAGTGGCTCCAGCAGTTCTGGC 225
Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 226 AGCTCCATTTCCAGTGGCCAGAGC----- 249
Qy 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 250 -----TCTGGTGGTGGCTCCAGTGGTAGCTCCAGCGGATCCAGCATTCGCCAGGGTGT 303
Qy 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 304 TCTGCA-----GGATCTTTTAAGCCAGGAACGGGGTATTCCAGGTC 345
Qy 90 AlaIle-----GlyGlyLysGlyPhePheGlnVal--- 99
Db 346 AGCTACTCTCCGGATCTGGCTAGTCTCTACAAGGTGATCCCGCTTCTCTCCAGCTGGGG 405

RESULT 31
LOCUS AY412935
DEFINITION Homo sapiens CDSN gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION AY412935
VERSION 1
KEYWORDS 1590 bp DNA linear GSS 16-DEC-2003
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1590)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
 Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1590)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
 Adams,M.D. and Cargill,M.
TITLE Direct Substitution
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
FEATURES Location/Qualifiers
 source 1..1590
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..>1590
 /gene="CDSN"
 /locus_tag="HCM4723"
ORIGIN
Alignment Scores:
Pred. No.: 0.003 Length: 1590
Score: 137.50 Matches: 113
Percent Similarity: 22.64% Conservative: 59
Best Local Similarity: 21.44% Mismatches: 185
Query Match: 5.32% Indels: 171
DB: 9 Gaps: 24

US-10-009-823A-1 (1-502) x AY412935 (1-1590)
Qy 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 169 ACTGG---AAGGGTGACTCCAGCGGCTTCAGTAGCTACAGTGGCTCCAGCAGTTCTGGC 225
Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 226 AGCTCCATTTCCAGTGGCCAGAGC----- 249
Qy 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 250 -----TCTGGTGGTGGCTCCAGTGGTAGCTCCAGCGGATCCAGCATTCGCCAGGGTGT 303
Qy 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 304 TCTGCA-----GGATCTTTTAAGCCAGGAACGGGGTATTCCAGGTC 345
Qy 90 AlaIle-----GlyGlyLysGlyPhePheGlnVal--- 99
Db 346 AGCTACTCTCCGGATCTGGCTAGTCTCTACAAGGTGATCCCGCTTCTCTCCAGCTGGGG 405

```

QY 100 -----ThrLeuAlaAspLysValHisTyr 107  
Db 406 AGCAGCAGCTCTCACTCGGGAAGCAGCGCTCTCACTCGGGAAGCAGCAGCTCTCATTCG 465  
QY 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127  
Db 466 AGCAGCAGCAGCAGCTTTTCAGTTTCAGTCAGCAGCAGCAGCTTCCAA----- 507  
QY 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro 147  
Db 508 GTAGGGAATGGCTCTGCTGCAACCAATGACACTCTTACCGCGGGAATCAACCTT 567  
QY 148 IleGln-LeuAsp-----PheAsnAspProThrValAlaLys 159  
Db 568 TCCAGCCTGGACAAAGCTTTCTTCCAAACCTTTGGGTATCCAGCAGTGGCCAA 627  
QY 159 sSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThr-- 178  
Db 628 ACCGTCAGCTCAACACGAGGTCCTGTA-----GTTCG-CACATCCCGACTCTCCCTG 680  
QY 179 -----AspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLys 196  
Db 681 CAGTGGAGGGCCATCGCTCGCACTCTGCGCCCTTACATCCAGCTCCCACTCTGTGTC 740  
QY 196 sGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgVa 216  
Db 741 AGGGGTCAGAGGCTGTGGTGGTG-----GT 767  
QY 216 lTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSe 236  
Db 768 GGTGACACGACGAGCTTCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821  
QY 236 rThrGly-----SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAs 253  
Db 822 CAATGGTGGCTTCCAGGCAAGCCCTGTCCTCCCAATCACTCTGTAGCAAACTCTATGG 881  
QY 253 pGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPh 273  
Db 882 TGGCTACGAGTGTGGTGGCTCTCTGACAGTATCTGTTCCAGGC---ATGACCTA 938  
QY 273 eSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293  
Db 939 CAGTAAGGGT-----AAAATCTATCTGCTGGCTTACTTCCACCA 977  
QY 293 sAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe----- 310  
Db 978 AGAG--AACCTGTGAAGGCTCTCA-----GGGTCCTCTCTTTCAGCTGG 1025  
QY 311 -----SerAl 312  
Db 1026 GCCCCCATCTCTGAGGCAAAATACTTCTCCAGCAACCCCATCATCCCCAGCAGTCGGC 1085  
QY 312 aAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGl 332  
Db 1086 AGCTTCTCGGCCATTCGTTCCAGCCAGTG-----GGGACTGTGGGTCCCA 1133  
QY 332 nAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLe 352  
Db 1134 GCTCTGTGGAGCGGCTCCAGCGGCTCCAGGACCTGCTCTCCCTCCAGTCTTCGAGT 1193  
QY 352 uProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSe 372  
Db 1194 CCCAGCAGTTCAGCATTTCCAGCAGCTCCGCTTCCACCTACCACTCCCTGGCGAGTGC 1253  
QY 372 rSerThrArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrI 392  
Db 1254 TTCCAGAGCCCTGCTCCACCCAGGC----- 1281  
QY 392 eThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAs 412  
Db 1282 -----ACCGCTCTTCCAGCAGCAGCTCC----- 1305

QY 412 nileProLeuAlaAArgPheThrSerGluAsp-----Gly 423  
Db 1306 -----AGTTCCAATCGAGTGCAGAAATCATCTTCAGCCTGTGG 1346  
QY 423 YLeuAArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSer-----Gly 439  
Db 1347 CAGCAAGTCCAGCTCTTCTGGTCACCTTCATGCTGCTCTCTCTCTGACACTGACTGG 1406  
QY 439 YGlyProGluPheGlyLeuPro-----GlyTh 448  
Db 1407 GGGCCCCGAT--GGTCTCCCACTCTGATCCCTCGCTGGTGCAGCCCTGTGGCTC 1463  
QY 448 rSerAsnTyrGlyLysLeu 454  
Db 1464 CAGCAGTGTGGAAGATC 1482  
RESULT 32  
LOCUS AY416869  
DEFINITION Pan troglodytes NUP153 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.  
ACCESSION AY416869  
VERSION AV416869.1 GI:39772829  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
REFERENCE 1 (bases 1 to 4428)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 4428)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.  
FEATURES  
source  
1..4428  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
gene  
1..4428  
/gene="NUP153"  
ORIGIN  
1..4428  
/locus\_tag="HCM6034"  
Alignment Scores:  
Pred. No.: 0.0244 Length: 4428  
Score: 135.50 Matches: 112  
Percent Similarity: 34.15% Conservative: 56  
Best Local Similarity: 22.76% Mismatches: 221  
Query Match: 5.24% Indels: 103  
DB: 9 Gaps: 21  
US-10-009-823A-1 (1-502) x AY416869 (1-4428)  
QY 16 SerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLys 35  
Db 3067 AGCTTTGGTACAGGTGTATTAACTCCACCCCTGCTCTTAACCATAGTACCTCT 3126  
QY 36 GlnGlnGlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGly 55  
Db 36

Db 3127 GAGAACAGAGCGCTTCAAC-----TTTGGACACACAGAA 3162  
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArg----- 72  
Db 3163 ACCAAGAGTCTTGTAGTGGCTCTTTCATGTGTAGACATCAGAAAGTAAAGAAAGAA 3222  
Qy 73 ThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGly 92  
Db 3223 ATGCCTGCCACCAAGAGGAGTCTCTTTTGGCAAC---GTGGAGCCTGCTCTCTGCCA 3279  
Qy 93 GlyIysGlyPhePheGlnVal-----ThrLeuGluAspIysValHisThrArgAla 110  
Db 3280 TGTGCTCAGTCTTGTGTGGAGGACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGTCT 3339  
Qy 111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer-----GlyPhe 127  
Db 3340 ACTTCTCTAGTGTGTGGAGAGGAGTCTCTTTTGGCAAC---GTGGAGCCTGCTCTCTGCCA 3379  
Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysIysGluThrLeuGluPro 147  
Db 3400 TCTTTGGGAATTCAGACAAACAAAGATGAGATCTTCAAAAGTCCACA----- 3450  
Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSer-----ProAlaLysThrSer 165  
Db 3451 TTTAGTCTTGTATGATCAAAACCATCTGAGAAGCAATCTGAACAGCCAGCAAAAGCCACT 3510  
Qy 166 ThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspIysThrGlnSerGluAla 195  
Db 3511 TTTGCTTTGGAGCTCAAACT-----AGTACTACAGCTGATCAAGGTGCAGCA 3558  
Qy 186 AsnProTyPheAlaLeuLeuGluSerTrpIysGlyAsnGlyThrProPheIle----- 203  
Db 3559 AAGCCAGCTTTTGTAGTTCTTGAACAAACAGTCTCTCTAGTTCACAGTACACCACTCTCT 3618  
Qy 204 -----SerThrSerAsnTySerTyAlaGlnProMetArg 215  
Db 3619 GCTGTGTGGGATATTTGTAGTCTTCCACTCTCTTCCCAATCCACTGTGGCTACCTTT 3678  
Qy 216 ValTyAspGlnGlnGlyAsnSerHisAspIleThrValTyPheAspGlyAlaProSer 235  
Db 3679 GTGTTGGACAGTCCAGCAATCCTGTGAGCAGCTCTGCTTGTGTAAACATCTGTAATCC 3738  
Qy 236 SerThrGly-----SerLysThrPheGluTy 244  
Db 3739 AGCACCTCTCAGCTCTTCTGTTTCTCAAGATAGCAAACTAGCAACCATCCAGCA 3798  
Qy 245 LeuValAlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer 262  
Db 3799 GGTACAGCTGTCAACCCATTTGCTCTTGTCTCAGGAGCCAGCAGTATATACTACCACC 3858  
Qy 263 AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeu----- 279  
Db 3859 TCTGTTTGGCTTTGGAGCCCAACCACTATCTAGCTCTGAGGATCTCTCTTTGTATTT 3918  
Qy 280 -----LysAsnMetThrAlaPhe-----ThrProThr 288  
Db 3919 GGAAGTGGACCTCAGCACCATCTGCTCCAGTCCAGCATTTGTGTCTAACCAGCCCAACA 3978  
Qy 289 GlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuPro 308  
Db 3979 TTTGGCAAAAGTCAA-----GTGTGCCAGCCAGCCCAATCCCGAGCTTTGGATCTATA 4032  
Qy 309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle 328  
Db 4033 TCATCTTCCAGCAGTATTTTCCCACTGGTTCTCAGCTGACCCACCTACTTTTGGACA 4092  
Qy 329 LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348  
Db 4093 GTGTCAAGCAGTGGCAG-----CCCCCTGTG----- 4119  
Qy 349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368  
Db 4120 TTTGACAGCAACCTAGT-----CAGTCTGCATTTGGCTCTGGAACTCTCT 4167

Qy 369 Asn-----GlySerSerSerThrArgArgTySerGlnAspGly 381  
Db 4168 AATCTTANNNNNNNTTCCAGTTTGGCAGCAGCAGCAGTCAAAATTCNCTTCAAAACAAC 4227  
Qy 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401  
Db 4228 AATCCATCAGGAGTGTTCACATTTGGTGTCAAAATTTAGCACACCTGCAGCCTCAGCCAG 4287  
Qy 402 TyrSerAsnSerGlnValValAspPheTyAsnIleProLeuAlaArgPheThrSerGlu 421  
Db 4288 CTTTCAGCTCGGGGCTTTCATTTAACAGTCTCCA---GCAGCATTTACA----- 4338  
Qy 422 AspGlyLeuArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyPro 441  
Db 4339 -----GTGGGCTCAAAATGGGAAAATGTCTCTCTCTCTCT----- 4374  
Qy 442 GluPheGlyLeuProGlyThrSerAsnTyGlyLys 453  
Db 4375 -----GGAACTTCAATCTCTGGTGGC 4395

RESULT 33  
CN823464  
LOCUS CN823464 726 bp mRNA linear EST 02-JUN-2004  
DEFINITION Oa splbn\_05M02\_M13reverse Sheep spleen\brain pSport1 library Ovis  
arles cDNA clone Oa\_splbn\_05M02\_5', mRNA sequence.  
ACCESSION CN823464  
VERSION CN823464.1 GI:47951533  
SOURCE EST.  
ORGANISM Ovis aries (sheep)  
Ovis aries  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Caprinae; Ovis

REFERENCE 1 (bases 1 to 726)  
AUTHORS Gosner, A. and Hopkins, J.  
TITLE Ovine spleen\brain cDNA library  
JOURNAL Unpublished (2004)  
COMMENT Contact: J Hopkins  
Veterinary Biomedical Sciences  
University of Edinburgh  
Summerhall Square, Edinburgh, EH9 1QH.  
Email: j.hopkins@ed.ac.uk  
Plate: 05 row: M column: 02  
Seq primer: M13reverse  
High quality sequence start: 15  
High quality sequence stop: 547.

FEATURES  
source  
1..726  
/organism="Ovis aries"  
/mol\_type="mRNA"  
/db\_xref="taxon:9940"  
/clone\_lib="Sheep spleen\brain pSport1 library"

ORIGIN  
Alignment Scores:  
Pred. No.: 0.00234 Length: 726  
Score: 133.50 Matches: 32  
Percent Similarity: 45.74% Conservative: 11  
Best Local Similarity: 34.04% Mismatches: 46  
Query Match: 5.16% Indels: 5  
DB: 7 Gaps: 1

US-10-009-823A-1 (1-502) x CN823464 (1-726)

Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23  
Db 432 TCTTTTAACTCGGCTCTTAGCGGCTCTATGCGGCAACAAACAACTGACGTCGCGC 491

Qy 24 AsnAsnIleAlaSerAlaAsnThrIleGlyTyLysGlnGlnValValPheGlnAsp 43  
Db 492 AACACATCGCACGTCGCGGACCAACCGGCTTCAATCGTCCCGCGCAATTTCTCGGAT 551



Db 4054 GGACAGCAGCTAGTGCAGTCTGCATTGGCTCTGGG----- 4089

Qy 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArg 376

Db 4090 -----ACAGCTAATGCCAGTTCGGTTTCCATTGGTAGCAGCACTAGCAATTC 4140

Qy 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGly 396

Db 4141 AACTTCACAAACAATAATCATCAGCAGGTGTTCCATTGGTGCAGTCCCGCAGCGCT 4200

Qy 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416

Db 4201 GCAGCTCAGCCGCGCTCAGTTCAGGGGCTTTTCATTAGCAGCACTCCA---GCA 4257

Qy 417 ArgPheThr 419

Db 4258 TCATTTTACA 4266

RESULT 35

CN762396/c

LOCUS

DEFINITION CN762396.1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA4AB01 5', mRNA sequence.

ACCESSION CN762396

VERSION CN762396.1 GI:47536319

KEYWORDS Acyrthosiphon pisum (pea aphid)

SOURCE Acyrthosiphon pisum

ORGANISM Acyrthosiphon pisum

REFERENCE 1 (bases 1 to 374)

AUTHORS Hunter, M., Martinez-Torres, D., Rahbe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Winkler, P.

TITLE An expressed sequence tags database for the pea aphid Acyrthosiphon pisum

JOURNAL Unpublished (2004)

COMMENT Contact: D. Tagu

INRA Rennes

UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.23.48.51.65

Fax: +33.2.23.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGGAACAGCTATGACC

Plate: 4

row: B

column: 1.

Location/Qualifiers

1. .374

/organism="Acyrthosiphon pisum"

/mol\_type="mRNA"

/cultivar="developmentstage"

/db\_xref="taxon:7029"

/clone="ID0AAA4AB01"

/tissue\_type="whole insect"

/dev\_stage="nymphs and adults (parthenogenetic females)"

/lab\_hosts="XLI-Blue"

/clone\_lib="ApMS"

/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University; Soil conditions: Soil; Sowing date: 01/06/1999; Harvesting date: 01/06/1999; Stress date: no stress; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction.; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

FEATURES

source

1. .374

/organism="Acyrthosiphon pisum"

/mol\_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN

Alignment Scores:

Pred. No.: 0.00119 Length: 374

Score: 132.00 Matches: 27

Percent Similarity: 62.50% Conservative: 13

Best Local Similarity: 42.19% Mismatches: 24

Query Match: 5.10% Indels: 0

DB: 7 Gaps: 0

US-10-009-823A-1 (1-502) x CN762396 (1-374)

Qy 438 SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsn 457

Db 367 TCTGGCGTGGCGCTGTGGGAGCAGCCGGGAGGAACTCTGGCTCCCTGACCACGGT 308

Qy 458 GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGln 477

Db 307 GCGTGTGTAGCGTCAACGTCGATCTCAGTAAGAACTGGTCAATATGATGTTGCCAG 248

Qy 478 ArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAla 497

Db 247 GGTACCTATCATGCTTAACGCCAGACCATCAAAACCCAGGACCGAGATCCTCAACACGCTG 188

Qy 498 LeuGluLeuLys 501

Db 187 GTTAACCTTACGC 176

RESULT 36

BG857482

LOCUS

DEFINITION 1024052806.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BG857482

VERSION BG857482.1 GI:14238666

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

REFERENCE 1 (bases 1 to 348)

AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2

JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. .348

/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

FEATURES

source

1. .348

/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN

Alignment Scores:



```
QY 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla 27
Db 216 GGCACAACTCAACATGGCGCAAAATAGTGGCTTTGGTACGACTAGTGGAGCAATT 275
QY 28 AsnAlaAsnThrIleGly-----Tyr 34
Db 276 GGAACATCTGCATTTGGTTCTAGCAACAATACTGGAGGCTTATTTGGAAATTCACAGACC 335
QY 35 LysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThr 54
Db 336 AAACAGGAGGATTAATTTGGTACCATTTAGCCAGCCAGCACTCCACAGCACT 395
QY 55 GlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 396 GGG-----TTTGGTTTGGCACATCAACAGGAATCAATCAATAGCTTA 437
QY 75 Phe-----ThrGlnGlyAlaPhe----- 80
Db 438 TTTGGAACTGCAAGTACCGGACCACTCTTTCTCATCCAGAACATGCAATTTGCACAA 497
QY 81 ---GluPro-----GlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db 498 AATAAACCAACTGGCTTTGGAAATTTGGNACCACTAGTACAGTGA-----GGATC 551
QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
Db 552 TTTGGAACATA-----AATACCACCTCTAATCTCTTTTGTAGACACA 593
QY 117 GluAspGlyPheLeuAsnAspProSerGlyPheThrLeu-----MetGlySerArgIle 134
Db 594 ---TCTGGCTCCCTTTTGGCCCAAGTAGTTTACAGCAGCACTACAGNACTACCATC 650
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 651 AAATTTAATCT----- 662
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 663 -----CCCATGGTACAGATAGTATGTCAAAGCTGAGTTAGCACT 704
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db 705 AACATCAGTACA-----AAGCATCAGTGTATTACTGCTGTAAGAA 746
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 747 TATGAAAGCAAGTCA-----TTAGAGGAACTA 773
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
Db 774 CGTTTGGAGGATTATCAGGCTAACCG-----AAGGCCACAG 812
QY 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
Db 813 AACCAAGTGGGAGGAGGACCCACCGCTTATTTGGTCTTCTCCAGCA----- 863
QY 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
Db 864 -----ACTTCCAGTGCAACAGGGCTTTCAGCTCCTCCACCACTAATCA 908
QY 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys-As 294
Db 909 GCCTTTTCATATGTCAGACAAATACTGCTTT-----GGAACCTAGCACAACTGGA 959
QY 294 PheAsnAlaTrpGlnProAlaProLeu-----ValAsnGlyLeuProGlnPheSe 311
Db 960 TTTGGAACAAATCCAGTGGTCTCTTTGGCCACAGAAATCAACAGTACCACTGCTCTTC 1019
QY 311 rAlaAsnPheValGlyAlaGlyIleGln-----ProLeuThrLe 324
Db 1020 AGCAAAACCAATTG-----GCAGGCTACACACCCCGAATATCGGCTTTCTTTGGTAAT 1076
QY 324 uAspPheGlyIleLysSerGlnGlnAsn---MetTrpAla----- 336
```

```
Db 1077 ACCAGCACCTTTGGACAGCCAGCAACCACTATTTGGCGCTATTTGGAGTAACCAAGCC 1136
QY 337 -----GlyAlaProAlaSerAlaIleAlaIleG 346
Db 1137 TCACAAACAGGAGTCTTTTGGACAGCTACAAACACACAGCAGCTGGGACAGCATTTGG 1196
QY 346 YThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerTh 366
Db 1197 GACAGGAACAGGCTCTTTTGGGAGCCCAATAGTGGATTTGGTGCATGTTGTCACCT 1256
QY 366 rAlaAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAs 386
Db 1257 GTTTGGCAATACAGCTTACAACTTTTGGAACCCAGCACACAGCTCTCTTCATTTGG 1316
QY 386 pLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerG 406
Db 1317 T-----ACACACAGTGGCGGCTCTTCCGGTTTGGCAACAAATAACAG 1358
QY 406 nValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
Db 1359 TGGAGCGATATTTTGGAGTAAGCCAGCAGCTGGAACTTTGGGAACCTGGACTT----- 1413
QY 426 gGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPr 446
Db 1414 ---GGTACAGGATTTGGACAGCTGCCTATGAAATCAGAGGTTTCATATTTCTCGACC 1469
QY 446 oGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMe 466
Db 1470 TGGAGTTACAGTACTT-GCAGCCACCATGATGGTGTGGA---TCAGTTGGATCT 1525
QY 466 tSerArgGlu 469
Db 1526 CCTGCAAGAG 1535
```

## RESULT 38

AZ577764/c

LOCUS

DEFINITION

SP: NGR234 genomic clone 14e10, genomic survey sequence.

ACCESSION

AZ577764

VERSION

AZ577764.1

SOURCE

GSS.

ORGANISM

Rhizobium sp. NGR234

Rhizobium sp. NGR234

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.

1. (bases 1 to 324)

Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.

Genetic snapshots of the Rhizobium species NGR234 genome

Genome Biol. 1 (6), RESEARCH0014 (2000)

21114532

PUBMED

COMMENT

Contact: Virginie Viprey

Laboratoire de Biologie Moleculaire des Plantes Superieures

University of Geneva

1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland

Tel: +44(0)1603450000

Fax: +44(0)1603450045

Email: virginie.viprey@bbsrc.ac.uk

Class: shogun.

Location/Qualifiers

1. .324

/organism="Rhizobium sp. NGR234"

/mol\_type="genomic DNA"

/strain="ANU265"

/db\_xref="taxon:394"

/clone="14e10"

/ANU265"

/notes="Vector: M13; derivative strain of NGR234 cured of

pNGR234a"

ORIGIN



```

QY 280 -----LysAsnMetThrAlaPhe-----ThrProThr 288
Db 3919 GGAACCTGGACCCCTCAGACATCTGCAGTCCAGCAATTGGTCTAACAGACCCCAACA 3978
QY 289 GlySerAlaThrLysAspLeuAsnAlaTpgInProAlaProLeuValAsnGlyLeuPro 308
Db 3979 TTTGGACAAAGTCAA-----GGTGCAGCAGCCCAATCCCGAGCTTTGGATCTATA 4032
QY 309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLysAspPheGlyIle 328
Db 4033 TCATCTTCCACAGCAATATTTCACCTGGTTCACGCTTCACACCTACTTTTGGACA 4092
QY 329 LysSerGlnGlnAsnMetTpaAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Db 4093 GTGTCAAGCAGTAGCCAG-----CCCTCTGTG----- 4119
QY 349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Db 4120 TTTGGACAGCAACCTAGT-----CAGTCTGCATTTGGCTCTGGAACAACCTCT 4167
QY 369 Asn-----GlySerSerSerThrArgArgTyrSerGlnAspGly 381
Db 4168 AATCTTANTTCGGGTTTCCAGTTTGCAGTGGTGGCAATCTAGCACACCTGCAGCCTCAGCCCA 4227
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 4228 AGTCCATCAGAGTGTTCACATTTGGTGCAATTTAGCACACCTGCAGCCTCAGCCCA 4287
QY 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 4288 CCTTCAGCTCGGGGGCTTTCATTTAACCAAGTCTCCA---GCAGCAATTACA----- 4338
QY 422 AspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyPro 441
Db 4339 -----GTGGGTCAATGGGAAATGTCTCTCTCTCT----- 4374
QY 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 4375 -----GGAACCTTCATCTCTGTCTCGC 4395

RESULT 40
B2577319 867 bp DNA linear GSS 17-DEC-2002
LOCUS msh2_5352.x1 msh Pseudomonas aeruginosa genomic clone msh2_5352,
DEFINITION genomic survey sequence.
ACCESSION B2577319
VERSION B2577319.1 GI:27212380
KEYWORDS Pseudomonas aeruginosa
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 867)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. 867
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5352"
FEATURES
source

```

```

/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 0.0119 Length: 867
Score: 128.00 Matches: 43
Percent Similarity: 43.75% Conservative: 20
Best Local Similarity: 29.86% Mismatches: 53
Query Match: 4.95% Indels: 28
Dbs: 8 Gaps: 5

```

US-10-009-823A-1 (1-502) x BZ577319 (1-867)

```

QY 11 GlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAlaAsn 30
Db 123 GGGGTGTGGCCGCCAGGACATTAACTGGCCACCATCTCCACAAACCTGGCCACGATATCC 182
QY 31 ThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeuAla 50
Db 183 ACCACCGCTTTCACGGCAACCGCGGTAGTTTCAGAAATATGGTGGACCAAGATTCGGGCC 242
QY 51 IleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln---ValGly 69
Db 243 CAGTCTGCGAGCTGTTCACCATGACAGCAGCTGTCTACGGGCTCCAGCTGGTCACC 302
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr--- 87
Db 303 GCGGGCGCGCTGTGCGCACCCAGAAATATCTATATCCCGGCAGCTGTAGACCAACCGAG 362
QY 88 -----AspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluAsp--- 103
Db 363 CAGCCGCTGGACATGGCAGCAACGCGTGGCTCTTCCAGGATCTCTGCTGCCGACCGT 422
QY 104 LysValHisTyrThrArgAlaGlyAsnPhe-----ArgPhe 115
Db 423 ACCGTGTTCTACACCTCTATGTGTGCTTCCACCTTGAACCTCCGACGGGCATATCGATAC 482
QY 116 ThrGln-----AspGlyPheLeuAsn 122
Db 483 TCTCAACTGCTTAGCCCTGGAGCCAGTATCGTGGTGCCCAACGAGACCCATCTTCAT 542
QY 123 AspProSerGly 126
Db 543 CGTCGGTCTGGA 554

```

Search completed: November 1, 2004, 04:12:54

Job time : 3058 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:16 ; Search time 40 Seconds  
(without alignments)  
1207.520 Million cell updates/sec

Title: US-10-009-823A-1  
Perfect score: 2586  
Sequence: 1 MMGSLFIGATGKTHSTGIG.....NSKSVTTADTMLQALELKR 502  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 557.5 | 21.6        | 428    | 2 AG0219 | flagellar hook pro |
| 2          | 534.5 | 20.7        | 463    | 2 C71292 | probable flagellar |
| 3          | 527.5 | 20.4        | 462    | 2 F83510 | flagellar hook pro |
| 4          | 509   | 19.7        | 442    | 2 C70135 | flagellar hook pro |
| 5          | 498   | 19.3        | 403    | 2 AB0640 | flagellar hook pro |
| 6          | 498   | 19.3        | 403    | 2 F10365 | flagellar hook pro |
| 7          | 493   | 19.1        | 401    | 2 F0810  | flagellar hook pro |
| 8          | 493   | 19.1        | 401    | 2 B85670 | flagellar biosynth |
| 9          | 491.5 | 19.0        | 402    | 2 A64851 | flagellar hook pro |
| 10         | 485   | 18.8        | 718    | 2 G71888 | flagellar hook pro |
| 11         | 484.5 | 18.7        | 411    | 2 A11161 | flagellar hook pro |
| 12         | 484.5 | 18.7        | 411    | 2 A11520 | flagellar hook pro |
| 13         | 481   | 18.6        | 718    | 2 F64628 | flagellar hook pro |
| 14         | 435   | 16.8        | 482    | 2 E70460 | flagellar hook pro |
| 15         | 426.5 | 16.5        | 545    | 2 F81420 | flagellar hook pro |
| 16         | 418   | 16.2        | 413    | 2 AE0089 | probable flagellar |
| 17         | 413.5 | 16.0        | 405    | 2 E84969 | flagellar hook pro |
| 18         | 388   | 15.0        | 434    | 2 D82106 | flagellar hook pro |
| 19         | 384.5 | 14.9        | 865    | 2 E81271 | flagellar hook cha |
| 20         | 380.5 | 14.7        | 591    | 2 B87361 | flagellar hook pro |
| 21         | 370.5 | 14.3        | 263    | 2 A83956 | flagellar hook pro |
| 22         | 367.5 | 14.2        | 425    | 2 A12646 | flagellar hook pro |
| 23         | 367.5 | 14.2        | 425    | 2 H97420 | flagellar hook pro |
| 24         | 354.5 | 13.7        | 605    | 2 D64623 | flag protein (AF32 |
| 25         | 349.5 | 13.5        | 605    | 2 G71881 | probable flagellar |
| 26         | 347   | 13.4        | 396    | 2 AE3529 | flagellar hook pro |
| 27         | 333   | 12.9        | 264    | 2 G69622 | flagellar hook pro |
| 28         | 325.5 | 12.6        | 261    | 2 B72243 | flagellar basal-bo |
| 29         | 311   | 12.0        | 262    | 2 C70372 | flagellar hook bas |

## ALIGNMENTS

## RESULT 1

AG0219  
flagellar hook protein FlgE [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C;Accession: AG0219

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Farrago, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; I  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, I

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AG0219

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <N>

C;Cross-references: UNIPROT:Q82PB5; GB:AL590842; PIDN:CAC90619.1; PID:G15979825; GSPDB:GN

C;Genetics:

A;Gene: flgE

| Query Match           | 21.6%            | Score 557.5;                                                     | DB 2;      | Length 428; |
|-----------------------|------------------|------------------------------------------------------------------|------------|-------------|
| Best Local Similarity | 29.1%            | Pred. No. 2.1e-27;                                               |            |             |
| Matches 145;          | Conservative 74; | Mismatches 196;                                                  | Indels 83; | Gaps 11;    |
| QY                    | 10               | TKMTHSTGLTGVSNINIANANTIGYKQOQVVFQDLFQSLAIGSTGSGNAGMGAQVG 69      |            |             |
| DB                    | 8                | SGMNAASNLVDVIGNNIANSATSGFKAGSVSFADMF-----GS-----QTGMGVKVA 55     |            |             |
| QY                    | 70               | SVRTFTQGAPEPGNSVTDLAIGKGFQVLTEDK---VHYTRAGNFRFTQDGFNDPSG 126     |            |             |
| DB                    | 56               | GTQDFNDGTATTNRRLDLAISONGFR--MQDSSGGIYARNGQFALDENRNVNMQG 113      |            |             |
| QY                    | 127              | FTLMGSRISNNPNKKTLEPIQLDNDPTVAKSPAKTALNAVNLGSDTKTQSEAN 186        |            |             |
| DB                    | 114              | LNLGYPATGTPPTVQGANPVLSPIDQWIS---AKATTSGNVANLTSTHVDIAEATS 170     |            |             |
| QY                    | 187              | PYFALLESKNGKNGPPISTSNYSYAQPMRVYDQGNHSDITV-YFDGAPSTGSKTEYL 245    |            |             |
| DB                    | 171              | PF-----DPNDPTYSFVNNMTFTSLGNHEINVFVYKRAEDATDGNWDVY 218            |            |             |
| QY                    | 246              | VAMNPSEDSASGSDTSAGLMSGTMTFSNGELKNWTAFTPTGSAFKDLNAWQAPLVN 305     |            |             |
| DB                    | 219              | TRDSSAKVTDPAEDTDPAAAKRGSMVDFDSNGALKVNTNGT-NATSTDTFTTIPMGVNV 277  |            |             |
| QY                    | 306              | GLP--QFSANFVGAGIQPLTDLDFGKQQQNMWAGAPASAAAIGTDIGKLSMMPIQISSG 363  |            |             |
| DB                    | 278              | GAPQSFALNVAGS-----KQGN-----KQGN----- 295                         |            |             |
| QY                    | 364              | NSTANGSSSTRYSQDGYPGQDLDVDTITSEGLQKQYNSQVDFYNTPLARFTSEDC 423      |            |             |
| DB                    | 296              | -----TGADSVIAVQNTGYAAGEFTGQINSQDGSVGTYSNQQTQLLQGIQVWVNFSPNEG 350 |            |             |

|    |       |      |     |   |        |                    |
|----|-------|------|-----|---|--------|--------------------|
| 30 | 307   | 11.9 | 261 | 2 | H83510 | flagellar basal-bo |
| 31 | 303   | 11.7 | 260 | 2 | H90810 | flagellar basal bo |
| 32 | 303   | 11.7 | 260 | 2 | D85670 | flagellar basal bo |
| 33 | 303   | 11.7 | 260 | 2 | C64851 | flagellar basal bo |
| 34 | 301   | 11.6 | 260 | 1 | XREBFG | flagellar basal bo |
| 35 | 301   | 11.6 | 260 | 2 | AD0640 | flagellar basal-bo |
| 36 | 291.5 | 11.3 | 423 | 2 | E97165 | flagellar hook pro |
| 37 | 290.5 | 11.2 | 262 | 2 | AC2644 | flagellar basal-bo |
| 38 | 290.5 | 11.2 | 262 | 2 | C97426 | flagellar basal-bo |
| 39 | 288   | 11.1 | 262 | 2 | F71801 | flagellar basal-bo |
| 40 | 287.5 | 11.1 | 454 | 2 | S61455 | flagellar hook pro |
| 41 | 287   | 11.1 | 264 | 2 | C71259 | probable flagellar |
| 42 | 286   | 11.1 | 262 | 2 | A64718 | flagellar basal bo |
| 43 | 285   | 11.0 | 260 | 2 | G84969 | flagellar basal-bo |
| 44 | 282.5 | 10.9 | 262 | 2 | AE3645 | flagellar basal-bo |
| 45 | 274   | 10.6 | 261 | 2 | AH0089 | probable flagellar |

QY 424 LRREGNHYATLDSGGPFLGPTSNYKLSVNOLETNSVDMSEMNMIHQGFQPM 483  
Db 351 LSEGNVWVKETQSSGNPTLTAGSGGFTLTSGALESNVDSLKELVNMVAQRYQSN 410  
QY 484 SKSVTTADTLMQALELK 501  
Db 411 AOTIKTDQILQTLVSLR 428  
RESULT 2  
C71292  
probable flagellar hook protein (flgE) - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: C71292  
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770; PMID:9665876  
A:Accession: C71292  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-463 <COL>  
A:Cross-references: UNIPROT:O07884; GB:AE001244; GB:AE000520; NID:G3323005; PIDN:AA65659  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0727

Query Match 20.7%; Score 534.5; DB 2; Length 463;  
Best Local Similarity 28.1%; Pred. No. 6.2e-26;  
Matches 154; Conservative 80; Mismatches 183; Indels 131; Gaps 14;  
QY 1 MMSLFGATGKTHSTGLGTVSNINANANTIGYKQGVVFDLFSQDLAIGS-----TG 55  
Db 1 MMSLFGVSGNQHQHTRDVIGNVANVTGFRGVNFQDLISQOLSAARNEEVG 60  
QY 56 SQGNQAGMGAVQSVRTITQGAFFPNSVTDLAIGKGFQVLTEDKVVHYTRAGNFR 115  
Db 61 GVNPKVEGLVLIASIDVHTQALQTTGINTDVSIOGSGFFVLKSEKTFTRAGAFV 120  
QY 116 TQDGLNDPS-----GFTLMSRLSNPNKKEITLPIQLDNDPTVAKSPAKTST 166  
Db 121 DNAGTLVNPANGVRQGMADVAGERLINSQAOTDLVIPG-----QKIDAQOTS 172  
QY 167 ALNAVNLGSDTKTQSEANPYFALLLESWKNGTPPSTNSYVAQPMRVYDQGNHSDI 226  
Db 173 TVHYACNLDKLPDLAADADEADVRSKW-----TTDF-----QYVDSFGQHTL 217  
QY 227 TVYFDGAPSSGKTFEYLVANP-----SDGSA-----ASGT 260  
Db 218 QINFSRVPGT--NNQWQATVAVDTEGTEVDQTRVGVTSDGAANTFVNFDFGLASVT 275  
QY 261 DSAGLLMSGTTFSNGELKXMTAFTPTGSATKDLNAWQAPLVNGLPQFSANFVCGAGI 320  
Db 276 DTAG-----NVTG--PTGV-----LLEASYDVUGA--- 299  
QY 321 PLTLDFGIKSOQNMWAGAPASAAIGTDIGKLPSPMPTQTSNGSTARN-----GSSST 374  
Db 300 -----NPDAGQVTRHAFTLNLGEI-----GTARNITQFAERSTT 335  
QY 375 RRYSDQGVPOGLVDVITITSEGLKQKYSNVQVDFYNIPLEPSTSEGLRREGNNHYSA 434  
Db 336 KATQDQYANGYLENFKIDQSGVITGVYNGVSDIGQLALAGFANQGGLEKAGENTYVQ 395  
QY 435 TLDSGGPFLGPTSNYKLSVNOLETNSVDMSEMNMIHQGFQMSKSVTTADTLM 494  
Db 396 SNNSGANISTGVMGKGLIAGTLENSVNDLTDQFTDMITQKGFQAGAKTITQSDTLM 455  
QY 495 QKALELKR 502  
Db 456 DTVLSLKR 463

RESULT 3  
F83510  
flagellar hook protein flgE PA1080 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: F83510  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathos  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83510  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-462 <STO>  
A:Cross-references: UNIPROT:O914P9; GB:AE004539; GB:AE004091; NID:G9946985; PIDN:AAG04465  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: flgE; PA1080  
Query Match 20.4%; Score 527.5; DB 2; Length 462;  
Best Local Similarity 28.5%; Pred. No. 1.7e-25;  
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;  
QY 4 SLFIGATGKTHSTGLGTVSNINANANTIGYKQGVVFDLFSQDLAIGTSGSQGPNQAG 63  
Db 2 SFGILSGIQAASSGLNVTGNINAGTVGFKQSAEFADYVAAV-LGS-----GSNPGQ 56  
QY 64 MGAQGVSVRTITQGAFFPNSVTDLAIGKGFQVLTEDKVVHYTRAGNFRPTDQGLND 123  
Db 57 SGLVLSDVYSQFKQGNIDSTNSVLDAINGFPFVTSNNGAISYTRAGYFNTDQDFVD 116  
QY 124 PGFTLMGSRISNNPNKKEITLPIQLD-FNDPTVAKSPAKTSTALNAVNLGSDTKTQ 182  
Db 117 NNGYRLQGYAVGPNQQLQGVVVDLKVERNQAPQATSSIQSYNLSNLTK-----P 168  
QY 183 SEANPYFALLLESWKNGTPPSTNSYVAQPMRVYDQGNHSDITVPE---DGAPSSP-- 237  
Db 169 PVTVPF-----DPSDAATYSSSLGIYDSQNSHTWSQFFIKNEPDPNATPP 216  
QY 238 ---GSKTFEYLV-ANPSEDSGAAGTDSAGLLMSGTTFSNGELKXMTAFTPTGS--- 290  
Db 217 IPENSWTMKVLIDGVNPLDPSNKP-----NSFNVTFDASQMTSVRA--PDGTTSG 266  
QY 291 -----ATKDLNAWQAPLVNGLPQFSANFVCGAGIPLTLDFGIKSOQNMWAGAPASAA 344  
Db 267 PGFSIDATNVIQSPA---TGNPPTP---GTGWIPAASD-GKTPPTYAWNGATGAASG 318  
QY 345 IGTDIGKLPSPMPTQTSNGSTARNSSSTRYS-----QDGYPOGLVDVITITSE 395  
Db 319 ISFDMRK-----TTQYSTAFQAQSNPIODGYTTGQLAGLEIDDT 356  
QY 396 KGLQKYSNVQVDFYNIPLEPSTSEGLRREGNNHYSATLDSGPGFGLPSTSNYKLS 455  
Db 357 GVIFARYTNGSKVQGVVLANFANIQLTPIGKTSWVQSSSEGEPAVGAPRSCTLGALQ 416  
QY 456 VNQLSTNVDMSEMNMIHQGFQMSKSVTTADTLMQALELK 501  
Db 417 SGALEASNDISNELVNLVHQRNYQNAKTIQEDAVTQTIINLR 462

RESULT 4  
C70135  
flagellar hook protein (flgE) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: C70135  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,  
son, D.; Peterson, J.; Kralavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997  
A;Authors: Smith, H.O.; Venter, J.C.  
A;Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.  
A;Reference number: A70100; MUID:98065943; PMID:9403685  
A;Accession: C70135  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-442 <KLE>  
A;Cross-references: UNIPROT:Q44767; GB:AB001137; GB:AB000783; NID:92688160; PIDN:AAC6666  
A;Experimental source: strain B31

Query Match 19.7%; Score 509; DB 2; Length 442;  
Best Local Similarity 28.2%; Pred. No. 2.3e-24;  
Matches 150; Conservative 71; Mismatches 190; Indels 120; Gaps 13;  
QY 1 MMSLSFGATGKTHSTGLTGVNNIANANTIGYKQOQVVFQDLFSDLAIGS-----TG 55  
DB 1 MMSLSYGVGLQNHQTMVDMVGNVNIANVNTIGFKKGRVNFQDMISQISGASRPDARG 60  
QY 56 SQGNQMGAGVGVRTITQAFPGNSVTDLAIGKGFQFVTLKDVHYTRAGNFRP 115  
DB 61 GTNPKQVGLGMNVASIDITHQAFPOSTQKASDLGVSGNGFFILKEGKNLFYTRAGFDV 120  
QY 116 TDGFLNDPGGFTLMGSRISN--NPNIKKTLEPIQLDFND---PTVAKSPAKTSTALNA 170  
DB 121 DSRHLNVAN---GNRIQHWARDLEGEKVINTASDIEDLIIPIDKXGAKSTKNVTF 176  
QY 171 VVNLGSTDKTQSEANPYFALLSWKNGTTPPISTNSYVAQPMRVYDQGNSHDITVYF 230  
DB 177 ACNLDKRLPLIQEGANPADIARTGWVNVKS-----LYDSFGN---VSVL- 217  
QY 231 DGAPSSGTSTFEVYLVAMNSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGS 290  
DB 218 -----ELR----- 220  
QY 291 ATKDLNA---WQAPLVNGLPQFSANFVGAGIPLTLDP-----GIKQQQNMWAG 337  
DB 221 VKDLNTPNLWNTVLING--EQNSNE-----TLGFDNEGALASLNGQPGQKGDILQ 270  
QY 338 APASAAAIGTDIGKLPSPMPTQSSG-----NSTARNSSSTRYSQDGYPGDLVDV 390  
DB 271 IPITFNVLGANVGEVGEQTVNLKLGTVGYSVTISITQFADSSSTKAIIDQYGMGWENY 330  
QY 391 TITSEGLQKYSNVQVDFYNIPLARFTSEGLRREGNNHYSATLDSGGPFGFLPQTSN 450  
DB 331 EIDQNGVIVGYNSGIRDLGKIALASFMPNPGGLAKSGDINFVETSSNGQVRIGETGLAG 390  
QY 451 YGKLSVNOLETNSVDMGEMVMIIIOGFQFMNSKSVTTADTMLOKALELK 501  
DB 391 LGDIRSGVLEMANVDLAEQFTDMIVTQGFQANAKTITTSQDLQLQELVRLK 441

RESULT 5  
AB0640  
flagellar hook protein FlgE [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium  
C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A;Note: This species has also been called *Salmonella typhi*  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AB0640  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar AB0502; MUID:21534947; PMID:11677608  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AB0640  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-403 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD08301.1; PID:916502347; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY1216

Query Match 19.3%; Score 498; DB 2; Length 403;  
Best Local Similarity 28.7%; Pred. No. 9.5e-24;  
Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;  
QY 10 TGMKTHSTGLTGVNNIANANTIGYKQOQVVFQDLFSDLAIGSTGSGPNQAGQAQVG 69  
DB 8 SGLNAATNLVDVGNVNIANSATYGFKSGTASFADWFA-----GS-----KVLGLGVKVA 55  
QY 70 SVRTIFTOGAFPGNSVTDLAIGKGFQFVTLKDVHYTRAGNFRFTQDGFNDPSGFT 128  
DB 56 GITQDFTDGTNTTNGRGLDVAISQNGFPFRLVDSNGSVFYSRNGQFKLDENLNMQGMQ 115  
QY 129 LMGSRISSNNPIKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTSSEANPY 188  
DB 116 LTGYPATGTPPTIQGANPAPITPTLMA---AKSITTASMQNLN-NSTDPVPSKT--- 168  
QY 189 FALLSWKNGTTPPISTNSYVAQPMRVYDQGNSHDITVYDGPASPSGTSTKTEYL 245  
DB 169 -----PFSVSDADSYNKKGTVTYVDSQGNADHNMVYF-----VKTQDNEWA 209  
QY 246 VAMNPSEDSAGTDSAGLLMSGTMTFSSNGELK-----NMTAFTPTGSKDNLNAWOP 300  
DB 210 VTTHDSSDPATAPTTA-----STLKENENGILESGETVNTTCTINGATA----- 256  
QY 301 APLVNLGPFQSFANFVGAGIPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLPSPMPTQ 360  
DB 257 -----ATFSLPLNS-----MQQN----- 271  
QY 361 SSGNSTARNGSSSTRYSQDGYPGDLVDVITSEGLQKYSNVQVDFYNIPLARFTS 420  
DB 272 GANNIVATN-----QNGYKPGDLVSYQINNDGTVVGNYSNEQEVQLGVLANFAN 322  
QY 421 EDGLRREGNNHYSATLDSGGPFGFLPQTSNYSKLSVNOLETNSVDMGEMVMIIIOGF 480  
DB 323 NEGLASQGDNVWAAQTQASGVALLGTAGSGNFGKLTNGALEASNVDSLSELVNMIVAQRY 382  
QY 481 QNNSKSVTTADTMLOKALELK 501  
DB 383 QSNAQTIKTQDQILNLTNLNLR 403

RESULT 6  
S10365  
flagellar hook protein flgE - *Salmonella typhimurium*  
C;Species: *Salmonella typhimurium*  
C;Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 09-Jul-2004  
C;Accession: S10365; S15357  
R;Homma, M.; DeRosier, D.J.; Macnab, R.M.  
J. Mol. Biol. 213, 819-832, 1990  
A;Title: Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their  
A;Reference number: S10361; MUID:90294298; PMID:2193164  
A;Accession: S10365  
A;Molecule type: DNA  
A;Residues: 1-403 <HOM>  
A;Cross-references: UNIPROT:P16322; EMBL:X51737; NID:G47671; PIDN:CAA36022.1; PID:947673  
R;Jones, C.J.; Macnab, R.M.; Okino, H.; Aizawa, S.I.  
J. Mol. Biol. 212, 377-387, 1990  
A;Title: Stoichiometric analysis of the flagellar hook-(basal-body) complex of *Salmonella*  
A;Reference number: S15353; MUID:90204563; PMID:2181149  
A;Accession: S15357  
A;Molecule type: protein  
A;Residues: 2-5 <JON>  
C;Genetics:  
A;Gene: flgE  
A;Map position: 23 min  
C;Keywords: basal body; flagellum  
F;2-403/Product: flagellar hook protein #status experimental <MAT>

Query Match 19.3%; Score 498; DB 2; Length 403;  
Best Local Similarity 28.7%; Pred. No. 9.5e-24;  
Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;



368 RNSSTRRYSQDGPQGLVDVITSEGLKQKYSNSQVDFYNIPLARFTSEDLRRE 427  
377 T-----TQNGYKPGDILVSYQINDDGTGVNGYNEQTQLLGGIVLFANFANNEGLASE 327  
428 GNNHYSATLSDGPECLPCTSNYKLSVNOLETNSVDMKSRMVMNMIILQRFQNMKSV 487  
328 GDNVWSATQSSGVALIGTAGTNGFGLTNGALASNDLSKELVNNVIVQRYQSNQRTI 387  
488 TTADTLMOKALELK 501  
388 KTQDQILNTLVNLR 401  
RESULT 9  
A:Accession: A64851  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: A64851  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: A64851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-402 <BLAT>  
A:Cross-references: UNIPROT:P75937; GB:AE000208; NID:G1787308; PIDN:AAC74160.  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: flgE; flak; fla FV  
C:Keywords: basal body; flagellum  
F:2-402/Product: flagellar hook protein flgE #status predicted <MAP>  
Query Match 19.8%; Score 491.5; DB 2; Length 402;  
Best Local Similarity 28.8%; Pred. No. 2.4e-23;  
Matches 143; Conservative 75; Mismatches 173; Indels 105; Gaps 15;  
10 TGMKTHSTGLTVSNNIANTIGYKQOVVFDLFSQDLAIGTSGOGFNOAGMAOVG 69  
8 SGLNAAATNLVDIGNNIANSATYGFKSGTASFAFMFA-----GS-----KVLGVKVA 55  
70 SVRTFTQGAPEPGNSVTDLAIGKGFQ--VTLEDKHYTRAGNFRFTQDGLNDPSGFT 128  
56 GITQDFTDGTNTTNGRLDVAISQNGFFRLVDSNGSVFYSRNGQFKLDENRNLVNMQLQ 115  
129 LMGSRISNNPIKETEPIQLDNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPY 188  
116 LTGYPATGTPPTIQGANPTNISIPNTLMA---AKTTTASMQNLN--NSSDPLPT-----166  
189 FALLESMKNGTPTPLISTSN---YSYAPQPMRYVDQGNSHDITVYFDGAPSTSGSKTPEYL 245  
167 -----VTFFSADNSYNKKGSVTVFDSQGNHADMVVF-----VKTGDNWQ-- 209  
246 VAMPNPSDGSAASTDSAGLMSGTMTFSSNGELKNNMTAFTPTGSAKDNLNMQAPLVN 305  
210 VVTQSDSPNSIAKT-----ATLEFNANGTL--VDGAMANNIATGAINGAEP-----256  
306 GLPQSANFVGAGIOPLTLDLGKISQONWAGAPASAAAIGTDIGKLPSPMPIOTSSGNS 365  
257 ---TFLSFLNS-----MQQN-----TGANNI 275  
366 TARNGSSSTRYSQDGPQGLVDVITSEGLKQKYSNSQVDFYNIPLARFTSEDLR 425  
276 VAT-----TQNGYKPGDILVSYQINDDGTGVNGYNEQTQLLGGIVLFANFANNEGLA 326  
426 REGNNHYSATLSDGPECLPCTSNYKLSVNOLETNSVDMKSRMVMNMIILQRFQNMKSV 485  
327 SEGDNVWSATQSSGVALIGTAGTNGFGLTNGALASNDLSKELVNNVIVQRYQSNQ 386  
486 SVTTADTLMOKALELK 501

387 TIKTQDQILNTLVNLR 402  
RESULT 10  
G71888  
flagellar hook protein - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: G71888  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: G71888  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-718 <ARN>  
A:Cross-references: UNIPROT:Q9ZKY0; GB:AE001509; GB:AE001439; NID:G4155350; PIDN:AAD06361  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: flgE  
Query Match 18.8%; Score 485; DB 2; Length 718;  
Best Local Similarity 25.0%; Pred. No. 1.4e-22;  
Matches 181; Conservative 94; Mismatches 222; Indels 226; Gaps 22;  
1 WMGSFIGATGKTHSTGLTVSNNIANTIGYKQOVVFDLFSQDLAIGTSGOGFNOAGMAOVG 56  
1 MURSLWSGVNGMAHQAIALDIENNIANNVNTGFKYASAFVDMLSQVKLATATAPYKXGL 60  
57 QGPN--QAGMGAQVGSVRTITQGAPEPGNSVTDLAIGKGFQFQVLTEDKV--HYTRAGN 112  
61 AGQNDPSVGLGVGDATTIKFSQGNQNTDKVTKLAIQDGFIIISPDRTGRNTRDGE 120  
113 RFTQDGLNDPSGGLTLM-----GSRISNNPIKETEPIQLDNDPTVAKSPAK 163  
121 FLFDSQGSVTTGGVLVQGVNRNGSDTKGKSDTDLKVDNTGPLENTRIDPMV-MEAR 179  
164 TGT--ALNAVNLGSDTKTQSEANPYFALLESWK--GNGTPISTSYSAQPMR-----215  
180 ASNRISMRANLAGRHADQTA---VFALDSSAKTSPDGINFVYDSGTNLQAQVAEDMGS 235  
216 VTDQGN-----SHDITVYFDGA-----PS-----STGSKTPEY 244  
236 LYNEGDGALLNENQGIWVSYKSAKVVDILPSAENSTLELNGVKISFTNDSAVRTSSL 295  
245 LVAMPNPSDGSAAAG-----TDSAGLLMSGTMTFSSNGELKN-----281  
296 VAKNAINAVKSGTGLEAVLDGKQLENTNLDGDEKLNIVVTQAGTGFANFLDGDK 355  
282 -MTAF-----TPTGSAKDNLNMQAPLVNGLPQFSANFV-----GAGIQPLTL 324  
356 DVTAIFYSYTHSISPNDIGQFRTTDLRALIHDANIVKDPISADNYQDSAAISGVTTIN 415  
325 DFG-----IKSQNNW-----AGAPASAA 343  
416 QYGMFEINKNKNVNIKENINIFVSGYSDSVTNVNVFKVAMKGLNTASLIEGASASS 475  
344 -----AIGT-----347  
476 KFTTHATSIDVIDSLGTKHAMRIEYRSGGAEMNFRVIVPEPGLVGGSAARNPVFEG 535  
348 -----DIGKPSMMP-----IQTSSGNSTARNSSSTRYS-----Q 379  
536 GRUHFVNDGSLAGMNPPLLQFPKNGADAPQRINLAFSGSGFDGLTSDVKLSETYAIQ 595  
380 DGYQDGLVDVITSEGLKQKYSNSQVDFYNIPLARFTSEDLRREGNNHYSATLSDG 439  
596 NGQAGDGLMDVRDSDGVLLGAFSGNRTLALAQVALANFANDAGLQALGCVFSGTNGSG 655  
440 GPBFGPLPGTSNYKLSVNOLETNSVDMKSRMVMNMIILQRFQNMKSVTTADTLMOKALE 499

Db 656 QALIGAANTGRGSGKLESSNDLSRLNLIWQGFQANSKAVTTSDQILNTLLN 715  
QY 500 LXR 502  
Db 716 LKQ 718  
RESULT 11  
AII161  
flagellar hook protein FlgE homolog lmo0697 [imported] - Listeria monocytogenes (strain  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AII161  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative Genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AII161  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <GLA>  
A:Cross-references: UNIPROT:Q92DV7; GB:NC\_003210; PIDN:CAC98775.1; PID:g16410086; GSPDB:  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0697

Query Match 18.7%; Score 484.5; DB 2; Length 411;  
Best Local Similarity 27.0%; Pred. No. 6.8e-23;  
Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 11;  
QY 1 MMGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVVFDLFSQD----LAIGSTGS 56  
Db 1 MNQMTYTAISGNVAFQQAALSVTSNNIANANTIGYKQSVFNDLLYQNTMGSVAGGLYAG 60  
QY 57 QGPNOAGMAQVGSVRTITFGAFEPGNSVTDLAIGGKGFQV--TLEDKVHYTRAGNFR 114  
Db 61 TNPMSFGSGSKIGAILTDYTAGSPTSGRNKDAALQGRFFIAGDNAGGNIVYTRDGSFA 120  
QY 115 FTQDGLNDPDSGFTLMGSRISNNENIKKTELPIOLDPNDPTVAKSPAKTSTALNAVNL 174  
Db 121 VSDNNILTTQQKVMGYATDKNGVNLNGNLQPIPLNSAIPGEATKNGSLSGNIPLDW 180  
QY 175 GDSTDKTQSEANPYFALLESMKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA- 233  
Db 181 GEK-DTISSE-----LSVYDNAGGKHLQVNMKAAT 210  
QY 234 PSSGSKTFFYLAVNPSDGSAASTDSAGLLMSGTMTFSSNGELKMTAFTPTGSAIK 293  
Db 211 PDASGNVSEYEIQM-----DGKALTPP-----VTGLTNAOQELTNPDA----- 251  
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTLDFGIKSQQNWAGAPASAAAIGTIDIGKLP 353  
Db 252 -LKNIQINSTVNG-----KQVNM-----GLNLSGLT 276  
QY 354 SMPITQSSGNSTARNSSSTRYSQDGYPOGDLVDVTTITSEKLGKQKYSNSQVDFYNI 413  
Db 277 NY-----GTNQVFSPTSDGKAATVKDYAVTDSGYIAVSYSDDGVIPVAQL 322  
QY 414 PLARFSDGLRRGNHYSATLDSGGPEGLPQTSNYGKLSVNOLETNSVNDMSREVMNM 473  
Db 323 AVATFSNEDGLVKMGNGEYVPGSSGDVAVYGAGNGAGGAGISGSSLEGNSVDLSREFVNL 382  
QY 474 IIRGQFQMSKSVTTADTLMQKALEL 500  
Db 383 MTYQSGFQGNKTVIRVADDDVMKQIVNL 409  
RESULT 12

AII1520  
flagellar hook protein FlgE homolog lin0705 [imported] - Listeria innocua (strain Clip11;  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AII1520  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative Genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AII1520  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <GLA>  
A:Cross-references: UNIPROT:Q92DV7; GB:AL592022; PIDN:CAC95937.1; PID:g16413157; GSPDB:GN  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin0705

Query Match 18.7%; Score 484.5; DB 2; Length 411;  
Best Local Similarity 27.0%; Pred. No. 6.8e-23;  
Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 11;  
QY 1 MMGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVVFDLFSQD----LAIGSTGS 56  
Db 1 MNQMTYTAISGNVAFQQAALSVTSNNIANANTIGYKQSVFNDLLYQNTMGSVAGGLYAG 60  
QY 57 QGPNOAGMAQVGSVRTITFGAFEPGNSVTDLAIGGKGFQV--TLEDKVHYTRAGNFR 114  
Db 61 TNPMSFGSGSKIGAILTDYTAGSPTSGRNKDAALQGRFFIAGDNAGGNIVYTRDGSFA 120  
QY 115 FTQDGLNDPDSGFTLMGSRISNNENIKKTELPIOLDPNDPTVAKSPAKTSTALNAVNL 174  
Db 121 VSDNNILTTQQKVMGYATDKNGVNLNGNLQPIPLNSAIPGEATKNGSLSGNIPLDW 180  
QY 175 GDSTDKTQSEANPYFALLESMKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA- 233  
Db 181 GEK-DTISSE-----LSVYDNAGGKHLQVNMKAAT 210  
QY 234 PSSGSKTFFYLAVNPSDGSAASTDSAGLLMSGTMTFSSNGELKMTAFTPTGSAIK 293  
Db 211 PDASGNVSEYEIQM-----DGKALTPP-----VTGLTNAOQELTNPDA----- 251  
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTLDFGIKSQQNWAGAPASAAAIGTIDIGKLP 353  
Db 252 -LKNIQINSTVNG-----KQVNM-----GLNLSGLT 276  
QY 354 SMPITQSSGNSTARNSSSTRYSQDGYPOGDLVDVTTITSEKLGKQKYSNSQVDFYNI 413  
Db 277 NY-----GTNQVFSPTSDGKAATVKDYAVTDSGYIAVSYSDDGVIPVAQL 322  
QY 414 PLARFSDGLRRGNHYSATLDSGGPEGLPQTSNYGKLSVNOLETNSVNDMSREVMNM 473  
Db 323 AVATFSNEDGLVKMGNGEYVPGSSGDVAVYGAGNGAGGAGISGSSLEGNSVDLSREFVNL 382  
QY 474 IIRGQFQMSKSVTTADTLMQKALEL 500  
Db 383 MTYQSGFQGNKTVIRVADDDVMKQIVNL 409

RESULT 13  
F64628  
flagellar hook protein flgE - Helicobacter pylori (strains CCUG 17874, NCT11637, 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
C:Accession: F64628; S78522; S61442; S61456  
R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodok, A.; McKenney  
son, J.D.; Kelley, J.K.; Cotton, M.D.; Weidman, J.N.; Fujii, C.; Bowman, C.; Watthey, L.  
Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; KarpX, P.D.; Smith, H.O.; Fraser, C.  
A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
A;Reference number: A64520; MUID:97394467; PMID:9252185  
A;Accession: F64628  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-718 <COM>  
A;Cross-references: UNIPROT:P50610; GB:AB000596; GB:AB000511; NID:G2313982; PIDN:AA0791  
B;O'Tool: P.W.; Kostrzynska, M.; Trust, T.J.  
submitted to the EMBL Data Library, May 1994  
A;Reference number: S78522  
A;Accession: S78522  
A;Molecule type: DNA  
A;Residues: 1-105; 'N', 107-236, 'Y', 238-258, 'P', 260-372, 'NS', 375-412, 'TI', 415-508, 'D', 510-  
A;Cross-references: EMBL:U09549; NID:G600723; PIDN:AAA92803.1; PID:G600725  
A;Experimental source: CCUG 17874; NCTC 11637  
B;O'Tool: P.W.; Kostrzynska, M.; Trust, T.J.  
Mol. Microbiol. 14, 691-703, 1994  
A;Title: Non-motile mutants of *Helicobacter pylori* and *Helicobacter mustelae* defective in  
A;Reference number: S61442; MUID:95198543; PMID:7891557  
A;Accession: S61442  
A;Molecule type: DNA  
A;Residues: 1-83; 85-105, 'N', 107-236, 'Y', 238-258, 'P', 260-372, 'NS', 375-412, 'TI', 415-508, 'D'  
A;Cross-references: EMBL:U09549  
A;Experimental source: CCUG 17874; NCTC 11637  
A;Note: only a part of the nucleic acid sequence is shown  
A;Accession: S61456  
A;Molecule type: protein  
A;Residues: 1-14, 'X', 16-25 <OZO>  
A;Experimental source: CCUG 915  
C;Genetics:  
A;Gene: flgB  
C;Keywords: flagellum

Query Match 18.6%; Score 481; DB 2; Length 718;  
Best Local Similarity 24.6%; Pred. No. 2.5e-22;  
Matches 178; Conservative 96; Mismatches 223; Indels 226; Gaps 21;

Qy 1 MMSGIFGATGKTHSTGLTVSNINANTIGYKQQVVPQDLFSQDLAIGS-----TGS 56  
Db 1 MLRLSWGVNGMAHQIALDIESNNIANVNTGFKYGRASFDVMSQVKLIATAPYKNGL 60

Qy 57 QGPN--QAGMAQGVSVRTIFTOGAFEPGNSVTDLAIGKGFFQVTDLKV--HYTRAGN 112  
Db 61 AGQDNFVGLGVGDVATTKIFSQNGI QNTDVKTDLAIGDGGFFIISPDGRITRNFIDGE 120

Qy 113 FRFTODGFLNDPDSGFTLM-----GSRISNNPNIKETLEPIQLDFNDPTVAKSPAK 163  
Db 121 FLFDQSGLVTTGGLVQGVWRNGSDTGNKGSDDTDLKVDNTGPLENIRIDPGMV-MPAR 179

Qy 164 TST--ALNAVNLGSDTDKTOSEANPYFALLESNK--GNGTPPISTNSYAO----- 212  
Db 180 ASNRISMPANLNGRHADQTA--IFALDSSAKTSDGINPYDVGTLNLAQVAEDMGS 235

Qy 213 -----PMRVYDQGG-----NSHDITVYPDGAP-----SSTGSKTFEY 244  
Db 236 LCNEDGDALLNENGIWVYSKAKWKDILPSAENSTLELNGVKISFTNDSAVSRSSSL 295

Qy 245 LVAMPNSEDGSAAG-----TTSAGLLMGSTWTFSSNGELKN----- 281  
Db 296 VAAKNAINAVKSQTEIAYLPGKQLRLNLTNKLDELKLNIVVQAGTAFANFLDGDK 355

Qy 282 -MTAF-----TPTGSAKDNLNAQWAPLVNGLPQFSANFV-----GAGIQPLTL 324  
Db 356 DVTAQKYSYTHSISPNADIGQFRITTEDLRALI QHDANIVKDPDLADNVQDSAAISGVSN 415

Qy 325 DFG-----IKSQQNMW-----AGAPASAA 343  
Db 416 QYGMPEINNKONKVNKLENLIFVSGYSDSVTNVLFNKNAMKGLNTASLIEGGASASS 475

Qy 344 -----AIGT----- 347  
Db 476 KETHATHATSDIVDSLGTGKHAIRIEFYRSGAENFRVIVPEPGLVGGSAARPVFE 535

Qy 348 -----DIGKLPMMMP-----IOTSSNGNSTARNSSSTRYS-----Q 379  
Db 536 GRLHENNDGSLAGNPPPLLQDPKNGADAPORINLAFGSSGFDGLTSDVDTYAEIQ 595

Qy 380 DCPQGDVLDVYTIISSEKLGKQKYSNSQVDFVFNPLARFTSEDDGLRRGNHYSATLDSG 439  
Db 596 NGYQAGDLMDFRFDGSDGVLGAFSNGRGTALAQAVALANFANDAGLQALGCVFSGTNSG 655

Qy 440 GPEFGLPGTNSYKLSVNLQLETSNVDMSREVMNIIQRFQMNSKSVTTADTLMLOKALE 499  
Db 656 QALIGAANTGRRGSISSGKLESSNVDLRSRLNLIVQRFQFQANSKAVTTSQILNTLIN 715

Qy 500 LKR 502  
Db 716 LKQ 718

RESULT 14  
E70460  
flagellar hook protein FlgE - Aquifex aeolicus  
C;Species: Aquifex aeolicus  
C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C;Accession: E70460  
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.  
Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
A;Reference number: A70300; MUID:98196666; PMID:9537320  
C;Accession: E70460  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-482 <AOE>  
A;Cross-references: UNIPROT:O67711; GB:AE000760; NID:G2984138; PIDN:AAC07675.1; PID:G2984  
A;Experimental source: strain VFS  
C;Genetics:  
A;Gene: flgE

Query Match 16.8%; Score 435; DB 2; Length 482;  
Best Local Similarity 27.6%; Pred. No. 1e-19;  
Matches 143; Conservative 94; Mismatches 204; Indels 78; Gaps 23;

Qy 1 MMSGIFGATGKTHSTGLTVSNINANTIGYKQQVVPQDLFSQDLAIG-----STGSQ 57  
Db 1 MLRSFYNAITGMDVSRFALDVTSDNLNANTVGFKKSRPIFQDMVSO-VVVGILNTTGT 59

Qy 58 GPNQAGMAQGVSVRTIFTOGAFEPGNSVTDLAIGKGFF--QVTDLKVHYTRAGNPRF 115  
Db 60 KTTTFGAGAVDSTOKVWTIGSFQTEITTDLAIGKALFILRDVLTNGYTYTRDGRFRI 119

Qy 116 TDGFLNDPDSGFTLMGSRISNNP--NIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 172  
Db 120 NREGYLINPGLYVQGVFKV--NPVTGEVGTQLEDIRVETQIPPKATG----EIVFNPT 173

Qy 173 NLGSDTDKTOSEANPYFALLESKNGTGPPISTNSYAOQPMRVYDQGNSHDITVYFDG 232  
Db 174 NLDERAPIIDQTTTP-----NPLDSFTYNYRYTLTIYDLSGREVPADIYF-- 219

Qy 233 APSSTGSKTFE-YLVA-----MNPSEDGSAASGSDSAGL-LMSGTMTFSSNGELKNM 284  
Db 220 --VKUTGNQWKVYFLASLXKERVINDWNGD-DDKTDIVFLDFNDQVHIADNGTFTSLPT 276

Qy 285 FPTGSGATKDLNAQWA--PLV-----NGLPQFSANFVAGIQLPLDPLFI-----K 333  
Db 277 F-----ASKTLE-FDPSTKGLVYIPGGDIVQDTAN-----QKFYLEVDLTPESGPSE 324

Qy 334 MWAGAPASAAATGTDIGKLPMMPIQTSSG--NSTARN-----GSSSTRYSQDGPQ 384  
Db 325 DPNDTSEYNKLGAKUGSETNKKIIVGEGILQNVNVIQNSYITQHAADFVITWDQGYAR 384

Qy 385 GDLVDVYTIISSEKLGKQKYSNSQVDFVFNPLARFTSEDDGLRRGNHYSATLDSGSGPF 443  
Db 385 GELIDLVLVLSGDSGVVGVYSGNETLPTRYRLAQAQTDPEELVKKGSNLYASVKT---PTI 441



A;Cross-references: GB:AP000398; GSPDB:GN00144  
A;Experimental source: strain APS  
C;Genetics:  
A;Gene: flgE; BU340

Query Match 16.0%; Score 413.5; DB 2; Length 405;  
Best Local Similarity 25.7%; Pred. No. 1.8e-18;  
Matches 131; Conservative 73; Mismatches 185; Indels 121; Gaps 13;  
QY 7 IGATGKTHSTGLGTGVSNNIANANTIGYKQOQVVFQDLFSQDLAIGSTGSGPNOAGMGA 66  
Db 2 IAIISGLLANNNDYMEIISNNIANASTIGYKSKPLFFDFMFSHSFVSNTTN-----GYGV 54  
QY 67 QVGSVRIFTQGAPEPGNSVTDLAIGKGPFPQ-VLEDKVHYTRAGNFRFTQDGLNDPS 125  
Db 55 GISSIIIONFNGMLVETGRDLGLIKDGFRLVDSQGHVYTRDQGFDFDKDQNIINIQ 114  
QY 126 GFTLMGSKRIS-----NNPNIKKETLEPIQDLDNDFTVAKSPAKTSTALNAVNLG--- 175  
Db 115 GMYITGLNTSCSKSDFNN-----RNLFPINL--KSNILKNKPTSEIMLKAFLNRTESK 168  
QY 176 ---DSTDKTQSEANPYFALLESWKNGTPTTSTSYAOPRMVYDQOQNSHDITVYFDG 232  
Db 169 SSVDSNDKLSKPEDYTIYS-----SFLTFNDGDELTSNDVFNITSFN- 205  
QY 233 APSSTGSKTFEYLAMNPSEDSASGSDSAGLLMSGTMTFSSNGELKNTAFTPTGSAT 292  
Db 206 -KKTNTKWTN--VESNDSDDKETKN-----SFLTFNDGDELTSNDVFNITSKDS 254  
QY 293 KDLNAQAPAPLVNGLPQFSANFVGAGIQPLTLDGFIKSGQQNMWAGAPASAAAGTIDGKL 352  
Db 255 KK-----YENITLNL----- 264  
QY 353 PSMPIQTSSNGNSTARNSSSTRYSQDGYPOGDLVDVITTSBKGKQKYSNSQVDFYN 412  
Db 265 -----TGTFBQNSDVSWEHSQNGYPOGKLTFTDIVTNGEIIIGTYCNQKQQTIGQ 315  
QY 413 IPLARFTSEDLREGNNHYSATLDSGGPEGL-PGTSNYKGLSVNOLETNSVDMREMV 471  
Db 316 ILLSKFINPEKLOPESGNLWSATAESGEAKTAMKAGIQESGVLNKTLEVSNDVNLNELI 375  
QY 472 NMIIIQRGFQNMKSVTATDMLQKALELK 501  
Db 376 NMIIAQRYQNSAQSFKEDIINTLINLQ 405  
RESULT 18  
Db2106  
flagellar hook protein FlgE VC2197 [imported] - Vibrio cholerae (strain N16961 serogroup  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C;Accession: D82106  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, W.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: D82106  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-434 <HEI>  
A;Cross-references: UNIPROT:Q9KQ10; GB:AE004292; GB:AE003852; NID:g9656760; PIDN:AAF9534  
A;Experimental source: serogroup O1; strain N16961; biotype B1 Tor  
C;Genetics:  
A;Gene: VC2197  
A;Map position: 1  
Query Match 15.0%; Score 388; DB 2; Length 434;  
Best Local Similarity 25.7%; Pred. No. 7.5e-17;  
Matches 132; Conservative 69; Mismatches 213; Indels 100; Gaps 17;  
QY 6 FIGATGKTHSTGLGTGVSNNIANANTIGYKQOQVVFQDLFSQDLAIGSTGSGPNOAGMG 65

Db 3 YVLSGLSAAQMDLNTTNNIANANTFGKSRAEFGDVISTSLF-----TNAKTTPQGG 57  
QY 66 AQVGSVRIFTQGAPEPGNSVTDLAIGKGPFPQVLEDKV-----HYTRAGNFRFTQDGL 121  
Db 58 VQAAKVAQOFHEGSSIIYNNPLDLRIAGTGFFAVAKDRLVPOQNELTRNGAFHLKNSFM 117  
QY 122 NDPQSFTLMGSRISNNP---NIKKTLEPIQL--DFNDPTVAKSPAKTSTALNAVNLGD 176  
Db 118 VTANDEFLGLEYV--NPDTGDVLSYEPKPINIPPOGKPE---KOTANIDLGANLSPAN-GD 171  
QY 177 STDKTQSEANPYFALLESWKNGTPTTSTSYAOPRMVYDQOQNSHDITVYFDGAPSS 236  
Db 172 LKDP-----ALFDI-----TDP---ETYNRTSTSIYDSMGOPKYLTYYLKDNQ 214  
QY 237 TGSKTFEYLAMNPSE-----DGSAASTDSAGLLMSGTMTFSSNGELKNTAFTPTGS 290  
Db 215 ANTWQTYTVDTKTEKPINVVGDAASPTGHVG---HTRFNNDGLTSS----- 261  
QY 291 ATKDLNANQP---APLVNGLPQFSANFVGAGIQPLTLDGFIKSGQQNMWAGAPASAAAGT 347  
Db 262 ---LNGQPIVTEPL-----GGANPVDLNL-----GA 285  
QY 348 DIGKLPMMPIQTSSNGNSTARNSSSTRYSQDGYPOGDLVDVITTSBKGKQKYSNSQV 407  
Db 286 DYNQTLSP-----SLDSATQPAAPFELTKFDQDGTATGTLTKIDFDENGSVLATYSNGIN 340  
QY 408 VDFYNIPLARFTSEDLREGNNHYSATLDSGGPEGLPGTSNYKGLSVNOLETNSVDM 467  
Db 341 TTLGRVALYRVANEQGLDKKGTQMDATQFSKAKTWGSGNKGSGFSISNGSLEQSNIDMT 400  
QY 468 REMANMIIIQRGFQNMKSVTATDMLQKALELK 501  
Db 401 QELVDLISQNRFOANSRALEVHNGIQQNILQIR 434  
RESULT 19  
E81271  
flagellar hook chain protein Cj1729c [imported] - Campylobacter jejuni (strain NCTC.11168  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: E81271  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrall  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: E81271  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-865 <PAR>  
A;Cross-references: UNIPROT:Q9PLU9; GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB7371;  
C;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: flgE2; Cj1729c  
Query Match 14.9%; Score 384.5; DB 2; Length 865;  
Best Local Similarity 19.9%; Pred. No. 3.4e-16;  
Matches 172; Conservative 100; Mismatches 229; Indels 365; Gaps 23;  
QY 1 MMSLFIGATGKTHSTGLGTGVSNNIANANTIGYKQOQVVFQDLFSQDLAIGST-----G 55  
Db 1 MMSRLMSGVSGQAQAVMDVGNLISNVTYTGFKYSRADFTGTFMSQTVKIATAPTDGKG 60  
QY 56 SQCPNOAGMGAOVGSVRTIFTQGAPEPGNSVTDLAIGKGPFPQVLEDKV--HYTRAGNF 113  
Db 61 GSNPLQIGLVSVSSTTRIHSQGSVQTTDKNTDVAINGDGFPMVSDGGLTNLYTRSGDF 120  
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKETLEPIQDLDNDFTVAKSPAKTSTA-----L 168  
Db 121 KLDAYGNFYNNAGFVVQGNINWDDQIDTSRTP-QNIFIDPGMHIPAAKSTEVAKANL 179  
QY 169 NAVNVLGDST-----DKTQ-----SEANPYFALL 192

Db 180 NSGLAIGTSRNLYALDSVHGWNKTQRAEDENDTGTTFYTTSKNSVEVTEKGVDAAGSL 239  
Qy 193 ESWKNG-----  
Db 240 FNAKQGLNLRDQGIWVSYADATYSTNKGVNAPDPNLOONQTAFAWGTANQVNLIDIT 299  
Qy 200 -----TPPISTSNYSYAQPMRVVDOOG----- 221  
Db 300 LINGVRIQNRADIOSIDDAIAYINTFTAPTDRDGTGKAVKNGKSGIDFVNDNADGITDN 359  
Qy 222 -----NSHDIITYPD-----GAP-----SSTGSK-----TF- 242  
Db 360 MKNINLVANTAGELMNAVNNNNQTFNNNGNGOAGTPTINKNGSSLTATNITFT 419  
Qy 243 -----EVLVAMN-----PSED 253  
Db 420 POPQOANTVQLTGGLNAQIITAHKIYSSNPVDIGPMYNDGPAFGPGANATRTPEP 479  
Qy 254 GSAAS-GTDSAGLLMSGTMTFSSNGELKXN----- 282  
Db 480 GSAAYWDAVNGGLLNTVTRFTEDLRELLQRDARYGVVDGSGTFAAADINQIKVVV 539  
Qy 283 -----TAPTPTGSATK-----DLNAWQ----- 299  
Db 540 TADGHFAISANEQSTVPPNAINGVGNATTTDPRKMSFNITAYSNKQGTVSTNDATFAIF 599  
Qy 300 ---PAPLVNG-----LPQFSANFVAGI-----QPLTLDFGIKQQN 333  
Db 600 KAFDGLVIGNQIKESQELKLSAFSAGLEIYDLSGSKHTLEVQVKQSTTQDGGNEQWMI 659  
Qy 334 MWAGAPAGAAIG-----TDIGKLPSPMP-----IOTSSGN 364  
Db 660 IRVPEPAEINTTGEPPNIIIVGTARFNDGSLASTYPTPTNFPNNGAAPNQIKLSFGT 719  
Qy 365 S-----TARNGSSTRYSODYPOGLVD--VTITSEGLQKGYNSQVVDYFNPLA 416  
Db 720 SGNNDGLVSSNSASLTQATDGYTSGNLKPDATRVDDKGNILGEFTNGKTFPAKATAMA 779  
Qy 417 RFTSDGLRRGNHYSATLDSGGPEFGLPGTSGNYKLSVNOLETNSVDMREMNMI 476  
Db 780 SVANNNGLEEIGNLFKVTANSNIVWGEAGTGGRGEMKTSALEMSNVDLRSITELIII 839  
Qy 477 QRGQFQWNSKSVTTADTLMQKALELKR 502  
Db 840 QRGYQANSKTISTSDMLQTLQIQKQ 865

RESULT 20  
B87361  
flagellar hook protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: B87361  
R.; Nierman, W.C.; DeBlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n. J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87361  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-591 <STO>  
A:Cross-references: UNIPROT:P35806; GB:AE005673; NID:gl3422168; PIDN:AAK22886.1; GSPDB:G  
C:Genetics:  
A:Superfamily: rod protein flag

Query Match 14.7%; Score 380.5; DB 2; Length 591;  
Best Local Similarity 22.1%; Pred. No. 3.4e-16;  
Matches 137; Conservative 98; Mismatches 230; Indels 155; Gaps 15;  
A:Gene: CC902

Qy 4 SLFTGATGKMTSTGLGTVSSNNIANANTIGYKQQQVVFQDLFSQDLAIGSTGSGPNQAG 63

Db 6 AMLAGVSLIANSSALAAISDNIANVNTVGFKRSTSNFSTLV-----TSCNKNQIYSA 58  
Qy 64 MGAQVGSVRTITFOGAFEPGNSVTDLAIGGKGF-----QVTLKDVHYTRAGNERFT 116  
Db 59 GGKQAQTHGFIQQGLTQSTTNLNDISISGAGFFVTEKPENLATDTRSFTRAGSFQDL 118  
Qy 117 QDGLNDPSGFTLMG-----SRISNNPIK 141  
Db 119 NLGFLRNDAGLYQLWGLADPVSLITPDSLDMLQLASINVSGVGTAEKTTVGVNANLR 178  
Qy 142 KETLEFIQDFNDPTVAKSPAKTISTALNAVNLG-DSDTKTQSEANPYFA---LLESWKG 197  
Db 179 SEQFVAAAAYSYKGT-AGSPSKTNVVDVSDATNHNHVDVYSTGIANPVSGNNEYLVDIKE 237  
Qy 198 NG---TPPI-----STSNYSYAQPM----- 214  
Db 238 NGVIVATGVAYDAATNELVSTIDYKGPVTSMTTTRINAAGTTVNLADLGIYNASG 297  
Qy 215 -----RVVD-----QQNSHDIITYFDGAPSTGSGTSTFEVLVAMNP-- 250  
Db 298 ADDAEVAVAGKLYDPSTWMSDYAKDMSKGVKDPFVQIPLSDSKGQRTVTLMLKGPQP 357  
Qy 251 -----SEDSASAGTDSAGLLMSGTMTFSSNGELKXMTAFTTGSATKDLNAWQAP 302  
Db 358 NQVYAELEKRAKPGDLAN--NNGQISTGILEFTIDGKLNKTSGLFTGTTPT-----AI 407  
Qy 303 LVNGLPQFSANFVAGIQPLTLDFGIKQQNMMWAGAPASAAAIGTDIGKLPSPMP 362  
Db 408 TIKSSGYIAPTTPPAVQPPT-----PPTWADALGIDQEV--QIDLASAA 451  
Qy 363 GNSTARNGSSSTRYSODYPOGLVDVITISEGLQKGYNSQVVDYFNPLAFTSED 422  
Db 452 GGLTQYNSQSVVQSVNTNGTAFGNLTNIEIDGGVSAIFDNGVTRRIAQVAIAFTSPN 511  
Qy 423 GLRRGNHYSATLDSGGPEFGLPGTSGNYKLSVNOLETNSVDMREMNMI 482  
Db 512 GLKGVNGNAYRVNTESSGTVSLKAPSGGAGALAPSTLEASTVDLSQFTGLITTRAYSA 571  
Qy 483 NKSQVTTADTLMQKALELKR 502  
Db 572 SSKITITADQMLEELINIKR 591

RESULT 21  
A83956  
flagellar hook protein flgE [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: A83956  
R.; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: A83956  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-263 <STO>  
A:Cross-references: UNIPROT:Q9KA41; GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA00616  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: flgE  
C:Superfamily: rod protein flgF

Query Match 14.3%; Score 370.5; DB 2; Length 263;  
Best Local Similarity 22.6%; Pred. No. 4.5e-16;  
Matches 115; Conservative 43; Mismatches 99; Indels 251; Gaps 9;  
A:Gene: flgE

Qy 1 MMSLDFIGATGKMTSTGLGTVSSNNIANANTIGYKQQQVVFQDLFSQDLAIGST----- 54  
Db 1 MLRSYSGISGRNRPQTGLDVIIGNNANVNTFGYKKGRTTFQDMVVSQOLA-GATAPGANR 59  
Qy 55 GSGQPNQAGMGAQVSVRTITFOGAFEPGNSVTDLAIGGKGFQVTLKDVHYTRAGNFR 114

```

Db 60 GGVPKQVGLGMLASIDNVHVGSGQNTNRELDLIGSGGFFQVQDGTFFYTSNGFY 119
QY 115 FTQDGLNDPDSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNL 174
Db 120 FDQGG-----TIVN- 128
QY 175 GDSTDXTQSEANPYFALLBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAP 234
Db 129 -----AQGLRLMDTNGN--PITV----- 144
QY 235 SSTGSKTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKXMTAFTTGSATKD 294
Db 145 -----PTAQSPSIGD-----GTVTV----- 161
QY 295 LNAWQAPLVNGLPQPSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAGTIDICKLPS 354
Db 162 ----- 161
QY 355 MPMIQTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQKYSNSQVVDFFYNIP 414
Db 162 -----IDEDGDLQ-----EAGMYR 175
QY 415 LAFTSEDGLRREGNNHYSATLDSGGPEFGLPQTSNYGKLSVNQLETSNVDMSSREVMNI 474
Db 176 LVNFAPEGLAKAGANLYLVANSAGPLFTSAGNGGTGDIAGTLEMSNVVD-SEEFTEMI 235
QY 475 IIOGFGQVMSKSVTTADTLMQALELKR 502
Db 236 VAORGQFQANTRIITSEILQELVNLKR 263

RESULT 22
AI2646
Flagellar hook protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AI2646
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AI2646
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <KUR>
A:Cross-references: UNIPROT:Q8UHV5; GB:AB008688; PIDN:AAL41591.1; PID:G17738927; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: flgF
A:Map position: circular chromosome

Query Match 14.2%; Score 367.5; DB 2; Length 425;
Best Local Similarity 26.2%; Pred. No. 1.4e-15;
Matches 135; Conservative 66; Mismatches 208; Indels 107; Gaps 13;

QY 1 MMGLFICATGCMKTHSTGLGTVSNNANANTIGYKQQQVWFQDLFSQDLAIGSTGSGPN 60
Db 3 IFGTWRTGVSGNNAQAKLGIVGDNANASTGYKRASTSFSSL-----VLPSSSGS---- 54
QY 61 QAGMGAQVGSVRTITFQGAPEFGNSVTDLAIGGKGFQVTLDEKVHY-TRAGNFRFTQDG 119
Db 55 YASGGVQSNVRSYISEQGNLSYTTSDTLAQNGGFFVVDGSGGTPPYLTRAGSFQKNSG 114
QY 120 FLNDPSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTD 179
Db 115 YLENAAGQLMGYPGSGNP-----PA-----AVN----- 139
QY 180 KTOSEANPYFALLEBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAPSTGS 239
Db 140 -----GFTGLEA-----INVNFG-----LTPASPSQGS 163
QY 240 KTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKXMTAFT-----T 286
Db 164 -----FPANLRDDKAATAP-----LPSGNSATAAFGNKTSLTAFDSCGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQPSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTTASPKSITID 271
QY 347 TIIGKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQKYSNSQ 406

```

```

Db 140 -----GFTGLEA-----INVNFG-----LTPASPSQGS 163
QY 240 KTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKXMTAFT-----T 286
Db 164 -----FPANLRDDKAATAP-----LPSGNSATAAFGNKTSLTAFDSCGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQPSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTTASPKSITID 271
QY 347 TIIGKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQKYSNSQ 406

RESULT 23
H97428
FigE protein (AF325192) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97428
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Boughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <KUR>
A:Cross-references: UNIPROT:Q8UHV5; GB:AB007869; PIDN:AAK86385.1; PID:G15155515; GSPDB:G
C:Genetics:
A:Gene: agrC_1012
A:Map position: circular chromosome

Query Match 14.2%; Score 367.5; DB 2; Length 425;
Best Local Similarity 26.2%; Pred. No. 1.4e-15;
Matches 135; Conservative 66; Mismatches 208; Indels 107; Gaps 13;

QY 1 MMGLFICATGCMKTHSTGLGTVSNNANANTIGYKQQQVWFQDLFSQDLAIGSTGSGPN 60
Db 3 IFGTWRTGVSGNNAQAKLGIVGDNANASTGYKRASTSFSSL-----VLPSSSGS---- 54
QY 61 QAGMGAQVGSVRTITFQGAPEFGNSVTDLAIGGKGFQVTLDEKVHY-TRAGNFRFTQDG 119
Db 55 YASGGVQSNVRSYISEQGNLSYTTSDTLAQNGGFFVVDGSGGTPPYLTRAGSFQKNSG 114
QY 120 FLNDPSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTD 179
Db 115 YLENAAGQLMGYPGSGNP-----PA-----AVN----- 139
QY 180 KTOSEANPYFALLEBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAPSTGS 239
Db 140 -----GFTGLEA-----INVNFG-----LTPASPSQGS 163
QY 240 KTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKXMTAFT-----T 286
Db 164 -----FPANLRDDKAATAP-----LPSGNSATAAFGNKTSLTAFDSCGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQPSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTTASPKSITID 271
QY 347 TIIGKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQKYSNSQ 406

```

Db 272 DVSGVQRAINLDLQMTQFSAKFTPGAVLNGGPSQ--IKDVEIGKDGGLVAVYQDGG 329  
QY 407 VVDFYNIPLARTSBDGLRENNHYSATLSDGGPEFGLPGTSGNYGKLSVNVQLETSDVM 466  
Db 330 RNNIYQALATVPVSDNLTPQNGVYLPNSDGVVITGFGQSGFGYIKQALGSGNVDI 389  
QY 467 SHWVNNYIIQGFQFOMNSKSVTTADTLMQKALELKR 502  
Db 390 ASELTDIESQRIYANSKVFQGSGLMDVLINLKR 425  
RESULT 24  
D64633  
Flagellar hook - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
C/Accession: D64633  
R/Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: D64633  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-605 <rom>  
A:Cross-references: UNIPROT:Q25566; GB:AE000600; GB:AE000511; NID:92314042; PIDN:AAD0795

Query Match 13.7%; Score 354.5; DB 2; Length 605;  
Best Local Similarity 23.5%; Pred. No. 1.5e-14;  
Matches 151; Conservative 82; Mismatches 222; Indels 177; Gaps 22;  
QY 1 MMGSLFIGATGKTHSTGLTVSNINANITGVKQQQVFDLFSQDLAIGTSGGPN 60  
Db 1 MNDTLNAYSIGIKTHQFGIDSLSNINANVTILGYRNDPEKTLFSSHLDAKSVAN 60  
QY 61 QAGGAQVGSVRTIFTO-GAPEPGNSVTDLAIGKGFF--OVTL-----EDK 104  
Db 61 DRNYGV-TGSGNVLNKGDEYMPGEFHMAYQKGWFIQGNKNGEITINKDGFQKQD 119  
QY 105 VHYTAGNFRFTQDGLNDPSGFTLMG-----SRISNNPNIKK---ETLPEIQL 150  
Db 120 NFLTRAGNFARDADGYLVTPEGYVYVYDGLKKIKDGLNSTARDEDIEKLHGLTSLPQI 179  
QY 151 -----DF-----NDP 155  
Db 180 PQDLTYQPVLSKTVNI SVNLNPKDHLKGVQDFLNDKGEIKERFLNQDINALANDNEP 239  
QY 156 TVAKSPAKTSTAL-----NAVNLGD-----STDKTQSEANPYFA 190  
Db 240 IDAITNRKLNISIQKEDGKEDFVFTYGDAGKGNQFKTLGDLQKLLKEXTGLDLN--- 295  
QY 191 LLESWKNGTTPPI-----STSNYSVAQPMRVYDQOG-NSHDITVYFDGAPSTSGSKTF 242  
Db 296 LIKSEKAKSPALLEIANPSETPIAFSLSGGIADKGLNANQWELKGISRDSVAIKIP 355  
QY 243 EYLVAMPSEDCSAAAGTDSAG---LLMSGTMTFSSNGELKNMTAFPTGSA--KOLNA 297  
Db 356 YV-----STEVDIYDKAGDKYLLQSEYMTNSN-----DPTSSPTSKRKNQT 397  
QY 298 WOPAPLV---NGLPQPSANFVGAGIQ-----PLTLDGFIKSOQNMWAGAPASAAA 344  
Db 398 WEVKSIVDPKNTPIINDPTWEIVGFSATHKMSAPMTLDF----- 439  
QY 345 IGTDIGKLPMMPIQTSNGSTARNSSSTRYSDGYPQGLVDVITITSEGLQKGYKN 404  
Db 440 ---KGNKLTSLDKSENHSDLSYQDSKLEASQDGKPRGIFRDMRIENGVSIFASN 496  
QY 405 SQWDFYNIPLARTSBDGLRENNHYSATLSDGGPEFGLPGTSGNYGKLSVNVQLETSDVM 453  
Db 497 GVPEVARIGLAFINDOGLKIGLNLYEMQEGTINGENRPLSGNPIGLWDEBKLKFGK 556  
QY 454 LSVNQLETSDGLRENNHYSATLSDGGPEFGL---SGPPEFGL---PQTSNYGKLSVNVQLE 502

Db 504 RIGILAFNTDQGRKIGKNLYEMQEGTINGENRPLSGNPIGLWDEBKLKFGKIRHKYLE 563  
QY 461 TSNVDMRSREVMNIIQGFQFOMNSKSVTTADTLMQKALELKR 502  
Db 564 TSNVAGNALTNLILMQGYSNWARAFGAGDDMIKEAISLKK 605  
RESULT 25  
G71881  
probable flagellar basal-body/rod/hookprotein - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C/Accession: G71881  
R/Alm, R.A.; Ling, D.S.L.; Moir, D.T.; King, B.L.; Brown, B.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: G71881  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-605 <arn>  
A:Cross-references: UNIPROT:Q9ZKUI; GB:AE001513; GB:AE001439; NID:94155407; PIDN:AAD0642;  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0844

Query Match 13.5%; Score 349.5; DB 2; Length 605;  
Best Local Similarity 22.8%; Pred. No. 3e-14;  
Matches 148; Conservative 86; Mismatches 224; Indels 191; Gaps 21;  
QY 1 MMGSLFIGATGKTHSTGLTVSNINANITGVKQQQVFDLFSQDLAIGTSGGPN 60  
Db 1 MNDTLNAYSIGIKTHQFGIDSLSNINANVTILGYRNDPEKTLFSSHLDAKSVAN 60  
QY 61 QAGGAQVGSVRTIFTO-GAPEPGNSVTDLAIGKGFFQV-----TLEDK 104  
Db 61 DRNYGV-TGSGNVLNKGDEYMPGEFHMAYQKGWFIQGNKNGEITINKDGFQKQD 119  
QY 105 VHYTAGNFRFTQDGLNDPSGFTLMG-----SRISNNPNIKK---ETLPEIQL 150  
Db 120 NFLTRAGNFARDADGYLVTPEGYVYVYDGLKKIKDGLNSTARDEDIEKLHGLTSLPQI 179  
QY 151 -----DF-----NDP 155  
Db 180 PQDLTYQPVLSKTVNI SVNLNPKDHLKGVQDFLNDKGEIKERFLNQDINALANDNEP 239  
QY 156 TVAKSPAKTSTAL-----NAVNLGD-----STDKTQSEANPYFA 190  
Db 240 IDAITNRKLNISIQKEDGKEDFVFTYGDAGKGNQFKTLGDLQKLLKEXTGLDLN--- 295  
QY 191 LLESWKNGTTPPI-----STSNYSVAQPMRVYDQOG-NSHDITVYFDGAPSTSGSKTF 242  
Db 296 LIKSEKAKSPALLEIANPSETPIAFSLSGGIADKGLNANQWELKGISRDSVAIKIP 355  
QY 243 EYLVAMPSEDCSAAAGTDSAG---LLMSGTMTFSSNGELKNMTAFPTGSA--KOLNA 297  
Db 356 YV-----STEVDIYDKAGDKYLLQSEYMTNSN-----DPTSSPTSKRKNQT 397  
QY 298 WOPAPLV---NGLPQPSANFVGAGIQ-----PLTLDGFIKSOQNMWAGAPASAAA 344  
Db 398 WEVKSIVDPKNTPIINDPTWEIVGFSATHKMSAPMTLDF----- 439  
QY 345 IGTDIGKLPMMPIQTSNGSTARNSSSTRYSDGYPQGLVDVITITSEGLQKGYKN 404  
Db 440 ---KGNKLTSLDKSENHSDLSYQDSKLEASQDGKPRGIFRDMRIENGVSIFASN 496  
QY 405 SQWDFYNIPLARTSBDGLRENNHYSATLSDGGPEFGLPGTSGNYGKLSVNVQLETSDVM 453  
Db 497 GVPEVARIGLAFINDOGLKIGLNLYEMQEGTINGENRPLSGNPIGLWDEBKLKFGK 556  
QY 454 LSVNQLETSDGLRENNHYSATLSDGGPEFGL---SGPPEFGL---PQTSNYGKLSVNVQLE 502

Db 557 IRHKYLETSVNVAGNALTNLILMORGYSNMARAFAGCDMDIKAI5LKK 605

RESULT 26

AE3529  
flagellar hook protein flgB [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C:Accession: AE3529  
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letevsa,  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD352; PMID:11756688  
A:Accession: AE3529  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-396 <KUR>  
A:Cross-references: UNIPROT:O8YDL6; GB:AE008918; PIDN:AAL53400.1; PID:gl7984295; GSPDB:G  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10159  
A:Map position: 11

Query Match 13.4%; Score 347; DB 2; Length 396;  
Best Local Similarity 25.68; Pred. No. 2.4e-14;  
Matches 130; Conservative 67; Mismatches 151; Indels 120; Gaps 15;

Qy 1 MMSLFTGATGMKTHSTGLGTVSNINANTIGYKQOOVFDLFSODLAIGSTGSGCPN 60  
Db 3 LYGMRTGVSGMAQAQNLSTVADNIANASTVGYKRAETQFSSLVL-----PS 50  
Qy 61 QAQNGAQQVGSVRT-----IFTQAFEPGNSVTDLAIGKGFFOVT-LEDKVHVTRAGNER 114  
Db 51 TAGO-YNSGSLVTVRVYGISDQGIKSTSTYDLDGNGYFVVGQGGSTYLTAGSFV 109  
Qy 115 FTQDGLNDPSGFTMGSRISNNPNIKKTELEPIQLDFNDPTVAKSPAKTSTALNAVNL 174  
Db 110 PDKNGDLVNSAGYLLGAGADEAAG---GLTVAGLNVNVNAAALPAEGSTAGDFTVNL 165  
Qy 175 GDSYDKTQSEANPYFALLEKSKNGTTPISTSNYSYAAPMEVYDQGNSHDITVYEDGAP 234  
Db 166 -PSTDQAP-----AAGGNHKTSLISYNDKGEKILTDFYF--- 199  
Qy 235 SSTGSKTPEYLVANMPSEDGSAAGTDSAGILMSGTMTFFSSNGELKNMTAFTPTGSATKD 294  
Db 200 TKTGAD-EWNVSVKNRADG-----VEIGTTLN-----FDPT---TGD 233  
Qy 295 LNAWQAPLVNGLQFQSFANFVGAGIQPLTDFGIKSQONMWAGAPASAAIGTDIGKLPS 354  
Db 234 -----LVSG-GNVAVNLGAYGGQTLNLL----- 256  
Qy 355 MNPOTSSGENTARNGSSSTRYSODGYPOGDLVDVTTITSEGLQKGYNSQVDFYNIP 414  
Db 257 -----GGSTQAGDYTIQAVINGQAPSSIKVDVNDGAVVAYVYENGTKQVLRIP 308  
Qy 415 LARFTSEDGLRRGNHYSATLDGSGPFGFLPGTSNYGKLSVNOLETNSVDMSEMNMI 474  
Db 309 LANVASPDRMTVVSQNIFLPSAESGDVRLGPPQGDGMGKIMSGTLENSNADIACELTDMI 368  
Qy 475 IIRGFQMNKSVITADTMDLQKALELKR 502  
Db 369 EAQRSYTANSVFQGFPELMDVLNLRK 396

RESULT 27

G59622  
flagellar hook protein flgB - Bacillus subtilis  
N:Alternate names: flagellar hook-basal body protein flgG  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: G69622; I42365; S14502; JG00233

QY 356 MFQIOTSSGNSTARNSSSTRYSQDGPQGLVDVDTITSEGLQKGYNSQVDFYNIPL 415  
Db 160 -----SIVD-----AEGTKQ-----DGGQIGI 176  
QY 416 ABFTSDGLRREGNNHYSATLDG-GPEFGLPGTNSYKLSVNOLETNSVNDMSREMNMI 474  
Db 177 VTFANSDGDKIGSNLYRESLNSGTASAAANOPGCGTGALKSGFLEMSNVDLTDEFTEMI 236  
QY 475 IIQRGFQMSKSVTTADTLMQKALELKR 502  
Db 237 VAQRGFQMSKSVTTADTLMQKALELKR 264  
RESULT 28  
B72243  
flagellar basal-body rod protein FlgG - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: B72243  
R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72243  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <ARN>  
A:Cross-references: UNIPROT:Q9X1M8; GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AAD366Q  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1542  
C:Superfamily: rod protein flgF  
Query Match 12.6%; Score 325.5; DB 2; Length 261;  
Best Local Similarity 22.4%; Pred. No. 2.8e-13;  
Matches 114; Conservative 37; Mismatches 103; Indels 255; Gaps 8;  
QY 1 MMSLSFIGATGKTHSTGLGTVSNNIANTIGYKQOVVFDLFSQDLAIGSTGSGPN 60  
Db 1 MMSLSIAATGSAQCFKLDITIANLANVDTGTGKYKRAEFQDLLYOYVKNAGTPTAATS 60  
QY 61 QAGMGAQVG-SVRT-----IFTQAFBPGNSVTDLAIGGKGFQVTTLED-KVHYTRAGNF 113  
Db 61 SLPTGLYVGHGVRTAAATRIIFLGNFQGTGALDLAGDGFQIQDGRYATREDGSF 120  
QY 114 RFTQDGLNDPSGFTLMGSRISNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173  
Db 121 KMDSEG-----RIVTSN----- 126  
QY 174 LGDSTDKTQSEANPYFALLESWKGNGTPTISTSNYSVAQPMRVYDQGNSHDITVYFDGA 233  
Db 127 -----RIVTSN----- 132  
QY 234 PSSTGSKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTTGSATK 293  
Db 133 -----GLLIVPEITI----- 142  
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKQQNMWAGAPASAAAIGTDIGKLP 353  
Db 143 -----PENAVSINV----- 151  
QY 354 SMIPIOTSSGNSTARNSSSTRYSQDGPQGLVDVDTITSEGLQKGYNSQVDFYNI 413  
Db 152 -----SPDGIVSAAELQDGTIQELG-----TI 172  
QY 414 PLABFTSDGLRREGNNHYSATLDGSGPERGLPCTSNYKLSVNOLETNSVNDMSREMNMI 473  
Db 173 TLVRFVNPGLKSGIDNLYIATPASGDFIEGVPGQDGFQKQGLFLEKSNVDVVRMVD 232  
QY 474 IIQRGFQMSKSVTTADTLMQKALELKR 502

Db 233 ITAQRAYEFNSRVIOQTADENMLRTATNVKR 261  
RESULT 29  
C70372  
flagellar hook basal body protein flgG - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C:Accession: C70372  
R.Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: C70372  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <AQF>  
A:Cross-references: UNIPROT:O67006; GB:AE000709; NID:g2983373; PIDN:AAC06963.1; PID:g2983  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: flgG2  
C:Function:  
A:Description: located in the distal rod  
C:Superfamily: rod protein flgF  
C:Keywords: Basal body; flagellum  
Query Match 12.0%; Score 311; DB 2; Length 262;  
Best Local Similarity 21.2%; Pred. No. 2.3e-12;  
Matches 108; Conservative 45; Mismatches 102; Indels 254; Gaps 9;  
QY 1 MMSLSFIGATGKTHSTGLGTVSNNIANTIGYKQOVVFDLFSQDL-----AIGSTGS 56  
Db 1 MPRALWTSAGMTAQTDLNLSHMANVNTVYKMKMATFQDLLYQVVEPSPGAPTPTT 60  
QY 57 QGFN--QAGMGAQVGSVRTIFTQGAFFPGNSVTDLAIGGKGFQVTTLED-KVHYTRAGNF 113  
Db 61 RSPSGFQIGLGYVSDTYGIFTQGNLTQDNLQDLIAIGDGFQVLPDGTIATYTRNGQF 120  
QY 114 RFTQDGLNDPSGFTLMGSRISNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173  
Db 121 RLDSEGRVNSDGYL-----DP----- 138  
QY 174 LGDSTDKTQSEANPYFALLESWKGNGTPTISTSNYSVAQPMRVYDQGNSHDITVYFDGA 233  
Db 139 -----EITL----- 142  
QY 234 PSSTGSKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTTGSATK 293  
Db 143 ----- 142  
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKQQNMWAGAPASAAAIGTDIGKLP 353  
Db 143 -----PADALSV----- 149  
QY 354 SMIPIOTSSGNSTARNSSSTRYSQDGPQGLVDVDTITSEGLQKGYNSQVDFYNI 413  
Db 150 -----SIAADGTVSVLR-----QGA-----TEVEEVGRI 173  
QY 414 PLABFTSDGLRREGNNHYSATLDGSGPERGLPCTSNYKLSVNOLETNSVNDMSREMNMI 473  
Db 174 ELAKFVNPAGLRRIGNNLYIQTATSGPFIIDNPGQGLTLLQGLYLESSNNVIVEMNVL 233  
QY 474 IIQRGFQMSKSVTTADTLMQKALELKR 502  
Db 234 IIAQRAYEFNTKIQAADEMLSQAANLRR 262  
RESULT 30  
H83510  
flagellar basal-body rod protein FlgG PA1082 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004



|    |     |                                                              |     |
|----|-----|--------------------------------------------------------------|-----|
| Db | 1   | MISSLIWIAKTLGDAOQTNDVIANNLANVSTNGFKQRAVFEDLLYQ--TIRQPQASSE   | 58  |
| QY | 61  | QAGM--GAQVGS-VRTIFT---QGAFEPGNSVTDLAIGKGFFOVTTLED-KVHVTRAG   | 111 |
| Db | 59  | QTTLSGQLIGTVRVPATERLHSGNLSQTNNSKVAIKGQGFQOMLPDSSAYTRDG       | 118 |
| QY | 112 | NFRFTQDGLNDPSGFTLMGSRISNNPNIKETLERIQLDNFNDPTVAKSPAKTSTALNAV  | 171 |
| Db | 119 | SFQVDONGQLVTAGGFQV-----QPAITIPA-----                         | 144 |
| QY | 172 | VNLGSTDKTQSEANPYFALLEWSKNGTPTPISNYSYQAQPMRVYDQCGNSHDTITVPD   | 231 |
| Db | 145 | -----NALSITIGRD                                              | 154 |
| QY | 232 | GAPSSSTGKTFEYLVAMNPSGDSGAASGTDAGLLXSGTWTFFSSNGELKNMTAFTPTGSA | 291 |
| Db | 155 | GVVSVT-----QOQQA                                             | 166 |
| QY | 292 | TKDLNAWQAPLVNGLPQFSANFVAGIQPLTLDFGIKSQCNMWAGAPASAAAIGTDIGK   | 351 |
| Db | 167 | -----                                                        | 166 |
| QY | 352 | LPSMMPIQTSSGNSTARNGSSSTRYSQDGYOQGLVDVITITSEGLQCKYNSQWDFY     | 411 |
| Db | 167 | -----PVQVGQLN-----                                           | 174 |
| QY | 412 | NTPLARFTSEGLRRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETSNVNMSREKV   | 471 |
| Db | 175 | ---LTTFNMDTGLESIGENLYTETQSSGAPNESTPGLNGAGLLYQGVWTSNVNVAEILV  | 231 |



Db 350 AYSNKLDPPTFISSAKEDNSGGYKQWGSALMSNVDLAQOQSDMIVASRAYQANGKIIT 409  
QY 489 TADTMLQKALELKR 502  
Db 410 TSDEILQDLVNLKR 423

RESULT 37  
AC2644  
Flagellar basal-body rod protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C/Accession: AC2644  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; exage, G.; Gilliet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, J.; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AC2644  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-262 <KUR>  
A/Cross-references: UNIPROT:Q44338; GB:AE008688; PIDN:AAL41569.1; PID:g17738903; GSPDB:G17738903  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: flgG  
A/Map position: circular chromosome  
A/Superfamily: rod protein flgG

Query Match 11.2%; Score 290.5; DB 2; Length 262;  
Best Local Similarity 21.5%; Pred. No. 4.3e-11;  
Matches 109; Conservative 44; Mismatches 86; Indels 267; Gaps 12;

QY 2 MGSIFGATGKTHSTGLTGVSNNTANNTIGYKQQVVDLPFSDLAIGTSGSQ-PN 60  
Db 1 MRALAAATGMDAQQTLEVIANNIANTTGYKRAAEFTDLYQ-----TERMGGVFN 55

QY 61 QA-----GMAQGVSVRTFTQG-AFEPGNSVTDLAIGKGFQVTLDE-KVHY 107  
Db 56 RANQAIPEGANIGLVQTSVAVRNIHTQNLITGNKL-DVAIIQGWFOIEAADGSTLY 114

QY 108 TRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQLDFNDPTVAKSPAKTSTA 167  
Db 115 SRAGAFNKADGNLVTDGVNVI-----PNI-----NIPT----- 144

QY 168 LNAVNLGDSDTKQSEANPYFALLESWKNGTTPISTNSYSAQPMRVYDQGNSHDIT 227  
Db 145 -----TRGQVTAIGNAAD--FTQLG-----DAQDIT 150

QY 228 VYFGAPSGSTGKTFEYLVAWNPSEDSAGSTDSAGLLMSGTMTFSSNGELKNMTAFTP 287  
Db 151 I----- 151

QY 288 TGSATKDLNAWQAPLVNGLPQPSANFVGAGIQPLTLDGFKSQONWAGAPASAAAIGT 347  
Db 152 ----- 151

QY 348 DIGKLPNMPIQTSSGNSSTANGSSSTRYSQDGYPOGDLVDVTTITSEGLKQKYSNSQV 407  
Db 152 -----TRGQVTAIGNAAD--FTQLG----- 171

QY 408 VDFYNIPLARTSDGLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMS 467  
Db 172 -----QLTIANFANRAGLPLGDNLFSTQPSAGPVPVGVDDPSYGYVKQSYLEGSNDVAV 227

QY 468 REMVNMIIQRGFQMSKSVTTADTM 493  
Db 228 KEITDLITAGRAYEMNSKVITTADEM 253

RESULT 39  
F71801  
Flagellar basal-body rod protein (distal rod protein) - Helicobacter pylori (strain J99)  
C/Species: Helicobacter pylori  
A/Variety: strain J99  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C/Accession: F71801  
R/Alm, R.A.; Ling, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999

RESULT 38  
C97426  
Flagellar basal-body rod protein flgG (distal rod protein) [imported] - Agrobacterium tumefaciens  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: C97426  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, R.; Goodner, B.; Wollam, C.; Allinger, D.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; A.; Liu, F.; Wollam, C.; Allinger, D.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: C97426  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-262 <KUR>  
A/Cross-references: UNIPROT:Q44338; GB:AE007869; PIDN:AAK86364.1; PID:g15155490; GSPDB:G15155490  
C/Genetics:  
A/Gene: AGR\_C\_972  
A/Map position: circular chromosome  
C/Superfamily: rod protein flgG

Query Match 11.2%; Score 290.5; DB 2; Length 262;  
Best Local Similarity 21.5%; Pred. No. 4.3e-11;  
Matches 109; Conservative 44; Mismatches 86; Indels 267; Gaps 12;

QY 2 MGSIFGATGKTHSTGLTGVSNNTANNTIGYKQQVVDLPFSDLAIGTSGSQ-PN 60  
Db 1 MRALAAATGMDAQQTLEVIANNIANTTGYKRAAEFTDLYQ-----TERMGGVFN 55

QY 61 QA-----GMAQGVSVRTFTQG-AFEPGNSVTDLAIGKGFQVTLDE-KVHY 107  
Db 56 RANQAIPEGANIGLVQTSVAVRNIHTQNLITGNKL-DVAIIQGWFOIEAADGSTLY 114

QY 108 TRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQLDFNDPTVAKSPAKTSTA 167  
Db 115 SRAGAFNKADGNLVTDGVNVI-----PNI-----NIPT----- 144

QY 168 LNAVNLGDSDTKQSEANPYFALLESWKNGTTPISTNSYSAQPMRVYDQGNSHDIT 227  
Db 145 -----DAQDIT 150

QY 228 VYFGAPSGSTGKTFEYLVAWNPSEDSAGSTDSAGLLMSGTMTFSSNGELKNMTAFTP 287  
Db 151 I----- 151

QY 288 TGSATKDLNAWQAPLVNGLPQPSANFVGAGIQPLTLDGFKSQONWAGAPASAAAIGT 347  
Db 152 ----- 151

QY 348 DIGKLPNMPIQTSSGNSSTANGSSSTRYSQDGYPOGDLVDVTTITSEGLKQKYSNSQV 407  
Db 152 -----TRGQVTAIGNAAD--FTQLG----- 171

QY 408 VDFYNIPLARTSDGLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMS 467  
Db 172 -----QLTIANFANRAGLPLGDNLFSTQPSAGPVPVGVDDPSYGYVKQSYLEGSNDVAV 227

QY 468 REMVNMIIQRGFQMSKSVTTADTM 493  
Db 228 KEITDLITAGRAYEMNSKVITTADEM 253

RESULT 39  
F71801  
Flagellar basal-body rod protein (distal rod protein) - Helicobacter pylori (strain J99)  
C/Species: Helicobacter pylori  
A/Variety: strain J99  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C/Accession: F71801  
R/Alm, R.A.; Ling, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999

A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A;Reference number: A71800; MUID:99120557; PMID:9923682  
A;Accession: F71801  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-262 <ARN>  
A;Cross-references: UNIPROT:Q9J26; GB:AE001570; GB:AE001439; NID:94156108; PIDN:AAD0706  
A;Experimental source: strain J99  
C;Genetics:  
A;Gene: flgG.2  
C;Superfamily: rod protein flgP

Query Match 11.1%; Score 288; DB 2; Length 262;  
Best Local Similarity 21.0%; Pred. No. 6.1e-11;  
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

Qy 1 MMSLFGATGKTHSTGLGTGVSNNANANTIGYKQQQVVFQDLFSDLAIGSTGSGQGN 60  
Db 1 MLRSLSYATSGMLAQOHTIITNNIANVNTTGFKSRADFNDLFYQACQYAGTNTSNT 60  
Qy 61 -----QAGGAGVGSVRTITQGAPEPGNSVTDLAIGGKGFQVILE-KVHYTRAGNF 113  
Db 61 LSPDGMVGLGVRPSAITMFSQSPKETENNDIAITGKGFQVQVLPDGTATYTRSGNF 120  
Qy 114 RFTQDGFNDPSTGLMGSRISSNNPNIKKTELEPIQLDFNDPTVAKSPAKTSTALNAVYN 173  
Db 121 KLDEQGLVTSSEVILLI-----PQI---TL-----PEDTTO-----VN 150  
Qy 174 LQSDTKTQSEANPYFALLESWKNGTPPISTNSYAPRMVYDQGNSHDITVYFDGA 233  
Db 151 IG-----VDTTQ-----VDGT 156  
Qy 234 PSSTGSKTFYLVAMNPSEDSRAASGTDLSAGLLSGTMTTFSSNGELKWTAFPTGSAIK 293  
Db 157 VSVT----- 160  
Qy 294 DLNAWAPAPLVNGLPQSFANFVGAGIQPLTLDGFIKQSNWAGAPASAAAIGTDICKLP 353  
Db 161 ----- 160  
Qy 354 SMPIQTSSNSTARNSSSTRYSQDGYPCQGLVDVITSEGLKQKYSNSQVDFYNI 413  
Db 161 -----QGLQTTSNVIG--QI 173  
Qy 414 PLAFPTSEDGLRRGNHYSATLDSGGPEFGLPQTSNYGKLSVNQLETSNVDSMRKVMN 473  
Db 174 TLANFVNPAGLHSGNDLFIITNASGDIAIVGNPDPSQGLKLRQGFLELSNVLVEENTDL 233  
Qy 474 IIIQRFQMSKSVTTTADTLMQKALELKR 502  
Db 234 ITAQRAYEANSKSIQTADAMLQTVNSLKR 262

RESULT 40  
S61455  
flagellar hook protein flgE - Helicobacter mustelae (fragment)  
C;Species: Helicobacter mustelae  
C;Date: 19-Mar-1997 #sequence\_revision 17-Jul-1998 #text\_change 18-Sep-1998  
C;Accession: S61455; S61444  
R;O'Toole, P.W.; Kosrzynska, M.; Trust, T.J.  
Mol. Microbiol. 14, 691-703, 1994  
A;Title: Non-motile mutants of Helicobacter pylori and Helicobacter mustelae defective  
A;Reference number: S61442; MUID:95198543; PMID:7891557  
A;Accession: S61455  
A;Molecule type: protein  
A;Residues: 1-41;420-454 <OTO>  
A;Experimental source: strain 4298  
A;Accession: S61444  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 31-39,'S',41-420 <OWO>  
A;Cross-references: EMBL:U09548  
A;Experimental source: strain 4298

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994  
C;Genetics:  
A;Gene: flgE  
C;Keywords: flagellum

Query Match 11.1%; Score 287.5; DB 2; Length 454;  
Best Local Similarity 25.7%; Pred. No. 1.5e-10;  
Matches 124; Conservative 59; Mismatches 169; Indels 131; Gaps 19;

Qy 1 MMSLFGATGKTHSTGLGTGVSNNANANTIGYKQQQVVFQDLFSDLAIGSTGSGQGN 54  
Db 1 MLRSLSYATSGMLAQOHTIITNNIANVNTTGFKSRADFNDLFYQACQYAGTNTSNT 60  
Qy 55 GSQGNQAGMGAGVGSVRTITQGAPEPGNSVTDLAIGGKGFQVILE-KVHYTRAGNF 112  
Db 61 GQNDVSVGLGVNNTTGVFSQSTQNTDVKTLAIEGDFGFIISPRGKTQNFTRDGE 120  
Qy 113 RFTQDGFNDPSTGLMGSRISSNNPNIKKTELEPIQLDFNDPTVAKSPAKTSTALNAVYN 163  
Db 121 KLDEQGLVTSSEVILLI-----PQI---TL-----PEDTTO-----VN 179  
Qy 164 TST--ALNAVNLGSDTKTQSEANPYFALLESWKNGTPPISTNSYAPRMVYDQGN 215  
Db 180 ASSNISLRANLNAGRHVDQV---AN-VFGLGTTKTPVDGINPIYDSDHNLTKQKADFGA 235  
Qy 216 VYDQGN-----SHDITVYFDGAPSTGSKTFEYLAVMNPSEDSRAASGTDLS 262  
Db 236 LFTQSGDAIGLTENOGIIVSVKTSWVNDIEATSGESSIEINHTRISFTNDSAAAGISSV 295  
Qy 263 -----AGLLSGTMTTFSSNG-----ELKN----- 281  
Db 296 VAAQNAINALKQKTGVEAFVNDGMLRLQKNMMDGDAEVKNIITADGTGAFANFIEGDS 355  
Qy 282 -MTAF---TPTGSATKDLNAWAPAPLVNGLPQSFANFVGAGIQPLTLDGFIKQSNWMA 336  
Db 356 DITAFRYRTTSGSPDSGTGQFRITEDLRALIQYDANLIKDPQSAIT----- 402  
Qy 337 GAPASAAAIGTDIGKLPNMPIQT-----SSGNSTARNSSSTRYS 378  
Db 403 ---DSTASVSVKFNKY-GMLEIQNKNDGDELKQDLNIFVSGYSSENST-NNGS-----E 451  
Qy 379 QDG 381  
Db 452 QDG 454

Search completed: October 26, 2004, 09:10:16  
Job time : 48 secs

